1 2 9 0

UNIVERSIDADE D
COIMBRA

João Alves Ferreira

Management of Hypertension Patients by means of remote monitoring solutions

Dissertação no âmbito do Mestrado em Engenharia Biomédica orientada pelo Professor Doutor Jorge Manuel Oliveira Henriques e pelo Professor Doutor Paulo Fernando Pereira de Carvalho apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Setembro de 2022

1 2 9 0
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

João Alves Ferreira

Management of Hypertension Patients by means of remote monitoring solutions

Supervisors:
Prof. Dr. Jorge Henriques (CISUC)
Prof. Dr. Paulo de Carvalho (CISUC)

Coimbra, 2022

This work was developped in collaboration with:

Centre for Informatics and Systems of the University of Coimbra

Altice Labs

Coimbra Hospital and Universitary Centre

Acknowledgments

First, I would like to sincerely thank my advisors, Prof. Jorge Henriques and Prof. Paulo de Carvalho, for all the guidance, support, and constant availability. All the assistance and knowledge imparted were essential throughout the development of this work.

To Prof. Marco Simões, for all his generous willingness to take the time to offer his expertise.

I am grateful to my parents, brother and the rest of my family for all the advice, understanding and patience over the years.

Finally, to all my friends, for all the companionship during the latest years and for celebrating even the smallest victories.

This work is funded by the project POWER (grant number POCI-01-0247-FEDER-070365), co-financed by the European Regional Development Fund (FEDER), through Portugal 2020 (PT2020), and by the Competitiveness and Internationalization Operational Programme (COMPETE 2020).

Abstract

Hypertension affects around 1.28 billion people and is one of the leading causes of cardiovascular disease, the major cause of death and disability worldwide. The poor prognosis of this disease is associated with high death and hospitalization rates. Studies on blood pressure have proved the importance of self-monitoring in hypertension control, treatment, and prevention. The development of digital solutions that promote lifestyle behaviors addressing the self-monitoring of blood pressure to manage the disease, improve compliance to treatment and achieve healthy living are therefore of major importance. In this sense, the present study focuses on the development of both a blood pressure prediction model and a knowledge-based recommendation system.

For the prediction module, four machine learning models for blood pressure prediction were developed and evaluated: a simple linear regression model, a long short-term memory (LSTM) neuronal network model, a jump neuronal network (JNN) model, and a case-based reasoning (CBR) model. Models were trained according to two prediction modalities, one focused on single day prediction (the last day of the prediction horizon) and another on multiple days prediction (all days of the prediction horizon). For each, different input data lengths and different prediction horizons were considered. All models were developed using data from the *MyHeart* study, which contains a blood pressure variable measured over 60 days for 41 patients. As for the recommendation module, a set of rules regarding changes in lifestyle habits for disease management and prevention was established by review of guidelines for hypertension control and treatment, issued by specialized cardiology institutions and societies. The rules were used as conditions in the development of a knowledge-based recommendation system to analyze patient information concerning factors such as exercise, diet, and alcohol consumption, and provide recommendations to improve these lifestyle habits, if necessary.

From the analysis of the results obtained from the prediction models, the JNN model was considered the most suitable for blood pressure prediction, having obtained the best performances in both modalities for all prediction horizons. Overall, the best result for this model was achieved for single day prediction, with a mean absolute percentage error of 3.64% and a root mean squared error and mean absolute error of 4.41 mmHg. These results are considered satisfactory for the purpose of this study, but due to the complexity of the problem, its practical application requires further analysis. Extending the prediction model by introducing other variables related to the evolution dynamics of blood pressure may be a promising approach to be explored. In the future, a clinical data collection study for this purpose will be carried out in collaboration with Altice Labs and CHUC, which will allow the validation and improvement of the already developed modules, both to be implemented concurrently in Altice's SmartAL remote monitoring solution for hypertension management.

Keywords: Hypertension, Blood pressure, Prediction, Knowledge-based recommendation system, Machine learning, Chronic diseases, Telemonitoring.

Resumo

A hipertensão afeta cerca de 1,28 mil milhões de pessoas e é uma das principais causas de doenças cardiovasculares, a principal causa de morte e incapacidade em todo o mundo. O mau prognóstico desta doença está associado a elevadas taxas de morte e hospitalização. Estudos acerca da pressão arterial provaram a importância da auto-monitorização no controlo, tratamento e prevenção da hipertensão. O desenvolvimento de soluções digitais que promovam estilos de vida orientados à auto-monitorização da pressão arterial, a fim de controlar a doença, melhorar a adesão ao tratamento e alcançar uma vida saudável, são, portanto, de grande importância. Neste sentido, o presente estudo centra-se no desenvolvimento tanto de um modelo de previsão da pressão arterial como de um sistema de recomendação baseado em conhecimento.

Para o módulo de previsão, foram desenvolvidos e avaliados quatro modelos de *machine learning* para a previsão de pressão arterial: um modelo simples de regressão linear, um modelo de rede neuronal *long short-term memory* (LSTM), um modelo de rede neuronal *jump* (JNN) e um modelo de raciocínio baseado em casos (CBR). Os modelos foram treinados segundo duas modalidades de previsão, uma focada na previsão de um único dia (o último do horizonte de previsão) e outra na previsão de vários dias (todos os dias do horizonte de previsão). Para cada uma, foram considerados diferentes tamanhos de dados de entrada e diferentes horizontes de previsão. Todos os modelos foram desenvolvidos com uso dos dados do estudo *MyHeart*, que contém uma variável de pressão arterial medida ao longo de 60 dias para 41 pacientes. Relativamente ao módulo de recomendação, foi estabelecido um conjunto de regras relativas a mudanças nos hábitos de vida para a gestão e prevenção da doença, através da revisão de diretrizes para controlo e tratamento de hipertensão, emitidas por instituições e sociedades especializadas em cardiologia. As regras foram utilizadas como condições no desenvolvimento de um sistema de recomendação baseado em conhecimento, com o intuito de analisar informações do paciente relativas a fatores como exercício, dieta e consumo de álcool, e fornecer recomendações no sentido de melhorar esses hábitos de vida, caso necessário.

Pela análise dos resultados dos modelos de previsão, o modelo de JNN foi considerado o mais apto para a previsão da pressão arterial, tendo obtido as melhores performances em ambas as modalidades para todos os horizontes de previsão. No geral, o melhor resultado para este modelo foi alcançado para a previsão de um único dia, com um erro percentual médio absoluto de 3,64% e um erro quadrático médio de raiz e erro absoluto médio de 4,41 mmHg. Estes resultados são considerados satisfatórios para o objetivo do presente estudo, mas, devido à complexidade do problema, a sua aplicação prática exige uma análise mais aprofundada. A extensão do modelo de previsão através da introdução de outras variáveis relacionadas com as dinâmicas de evolução da pressão arterial poderá ser uma abordagem profícua a explorar posteriormente. No futuro, realizar-se-á um estudo de recolha de dados clínicos para este fim, em colaboração com a Altice Labs e o CHUC, que permitirá a validação e melhoria de ambos os módulos já desenvolvidos e a sua implementação em paralelo na solução de monitorização remota da SmartAL da Altice para a gestão da hipertensão.

Palavras-chave: Hipertensão, Pressão arterial, Previsão, Sistema de recomendação baseado em conhecimento, Machine learning, Doenças crónicas, Telemonitorização.

Contents

List of Figures

List of Tables

List of Abbreviations

1

Introduction

1.1 Motivation

Hypertension refers to a higher than normal blood pressure in the arteries. Known as the "silent killer", hypertension affects over 150 million people across Europe and more than 1 billion throughout the world, and it is considered one of the leading causes of cardiovascular disease, the major cause of death worldwide [6, 7]. The high death and hospitalization rates associated with heart problems are mainly due to their poor prognosis and that is why telemonitoring could help executing the needed therapy in hypertension management by involving the patient in their own treatment [8]. Studies on Blood Pressure (BP) have suggested that its self-measurement can be a more efficient approach in BP control than when it comes to the care only based on office measurement [9–11]. Therefore, self-monitoring of BP is an increasingly important part of hypertension management, since Home Blood Pressure (HBP) is a better predictor of hypertension events than standard office measurements and self-monitoring strategies are easily introduced into the patient's routine [12]. For that matter, the development of digital solutions, like remote monitoring systems, aimed at enhancing lifestyle behaviors addressing the BP and disease management are of utmost importance in achieving healthy living.

1.2 Context

The project

The present thesis is integrated in the "POWER - Empowering a digital future" project, that aims to design innovative cloud and cognitive technologies based products and services, aligning with multiple strands of technological transformation: 5G networks, Edge / Cloud computing continuum, data-driven technologies and business models, and Artificial Intelligence. Being divided into five sub-projects, this thesis fits into sub-project 4 - "Future Services", in which Altice labs, alongside with the Centre for Informatics and Systems of the University of Coimbra (CISUC) and the Coimbra Hospital and Universitary Centre (CHUC), aim to the development of platforms for the analysis and prediction of physiological signals, and their use in the management of chronic diseases, such as, in this particular case, hypertension.

Main goal

The main goal of the development of a remote management system for hypertension patients based on BP prediction is to create a model that can anticipate a significant raise or reduction in one's blood pressure and act on it, ultimately in order to enable the early detection of a hypertensive or hypotensive crisis. By means of remote monitoring solutions, the clinical data collected by the self-monitoring system introduced into the patient's routine will allow the analysis and supervision of their condition and the autonomous establishment of BP-lowering habits in their lifestyle.

Challenges

Various barriers may still be inhibiting a broadened implementation of telemonitoring in management of hypertension. One of them may be related to the lack of quality and easily accessible databases containing relevant contextual data, that can provide a broader insight on patients' conditions and risk of developing hypertension. Even though some databases exist, the majority is not publicly accessible.

Additionally, in order to build tailored and patient-specific systems, each patient needs to be profiled, which could require a significant amount of background information about the patient's clinical history (e.g., stage of the disease and other comorbidities). This can be a complex task that might have to involve healthcare providers at an early stage. Also at the patient level, it is still a reality that some people have fewer technological skills or unreliable equipment and internet access, which is inconvenient for approaches where data entries are critical [13].

Proper data collection for the creation of the necessary databases and the establishment of certain implementation protocols and user guidelines may boost the establishment of telemonitoring systems for hypertension in healthcare.

Methodology

This thesis' work can essentially be divided into two components: the prediction module and the recommendation module. The former concerns the development and testing of different algorithms and models in the context of blood pressure prediction and the latter covers the recommendation model responsible for the patients' lifestyle optimization in order to maintain healthy BP values.

Blood pressure is the most fundamental aspect to consider when monitoring hypertension patients, since it is directly correlated with the disease. It describes the arteries' resistance to blood flow and is as high as the more blood is pumped by the heart and the narrower the arteries. Hypertension is known to be one of the leading causes of several cardiovascular and renal diseases and it is caused by a significant raise in blood pressure over time, being the first grade of hypertension defined by systolic and diastolic BP values over 140 and 90, respectively [6].

The variation and tendency of BP values are defined by a time series problem and, therefore, multiple machine learning and deep learning algorithms are found useful for its analysis and forecast. To solve this problem, BP prediction algorithms must be trained, making use of databases containing the necessary data. The performance of the existing algorithms is then assessed using evaluation metrics, and the overall best performing model is selected as the prediction algorithm to be used along with the recommendation model.

The recommendation module is designed to advise patients into optimizing their own life habits (e.g., physical exercise, dietary habits, weight control, etc.). A knowledge-based recommendation system is used to address this issue, relying on professional help from experts at Coimbra Hospital and Universitary Centre (CHUC) for its research and implementation.

1.3 Goals and Contributions

Besides the main goal, which is to research and implement a set of models to improve hypertension patients' lives by supporting remote monitoring systems, the expected contributions of this thesis are as follows:

- Specializing an already existing Case-Based Reasoning (CBR) framework, developed by the project's research team, for blood pressure prediction.
- Development of other models for the same end, namely using deep learning algorithms.
- Research on information that allows the creation of a set of rules on lifestyle habits (related to diet, exercise, etc.), that can be suggested to patients in the context of a knowledge-based recommendation system.
- Data collection study performed in collaboration with the Cardiology Department of CHUC, to complement the data currently available to the team on private databases.

- Support the studies related to diabetes taking place in the framework of the POWER project.

1.4 Structure

This document contains five chapters beyond the introduction.

Chapter 2 presents background information related to multiple concepts that will be referred throughout this document, such as the cardiovascular system, blood pressure, hypertension, and time series and its prediction.

Chapter 3 presents the state of art concerning telemonitoring, the blood pressure prediction module, and the recommendation module in this context.

Chapter 4 describes the methodology proposed to achieve the mentioned objectives, of both the blood pressure prediction algorithm and the recommendation system.

Chapter 5 contains the results obtained for the application of the proposed methodology and their discussion taking into consideration the context of the study.

Chapter 6 presents a conclusion to this study and future work.

Finally, appendix A contains the application programming interface (API) of the prediction module and B the API of the recommendation module.

2

Background

This chapter introduces the basic concepts needed to understand this document. Section 2.1 presents an overall view on the cardiovascular system and its constituents. Section 2.2 presents an explanation on blood pressure, as well as physiological and non-physiological factors that influence it. Section 2.3 adds an objective view of hypertension and its distinct types. Finally, section 2.4 introduces a background on time series prediction and algorithms used to address this matter.

2.1 The Cardiovascular system

The cardiovascular system is a sophisticated system in the human organism that is necessary to meet cells' demands for exchanges with the environment. There are three components of the cardiovascular system: the heart, blood vessels, and even though it doesn't contain blood, the lymphatic system, that allows important exchanges to happen between itself and blood vessels. The heart pumps blood into two circulations: the pulmonary circulation and the systemic circulation. The pulmonary circulation concerns the blood flow through the lungs and is involved in the exchange of gases between the blood and the alveoli. The pulmonary circulation is responsible for carrying venous (deoxygenated) blood to the lungs, returning arterial (oxygenated) blood to the heart, whereas the systemic circulation involves all remaining blood vessels that are not located in the lungs. The right side of the heart is responsible for receiving, through the superior and inferior vena cava, venous blood from the systemic circulation and pumping it, through the pulmonary artery, into the pulmonary circulation. After the gas exchanges in the lungs, arterial blood returns to the left side of the heart, by means of the pulmonary veins, and is pumped again into the systemic circulation, through the aorta, reaching all the remaining tissues [1]. Figure 2.1 presents a schematic representation of the cardiovascular circulation.

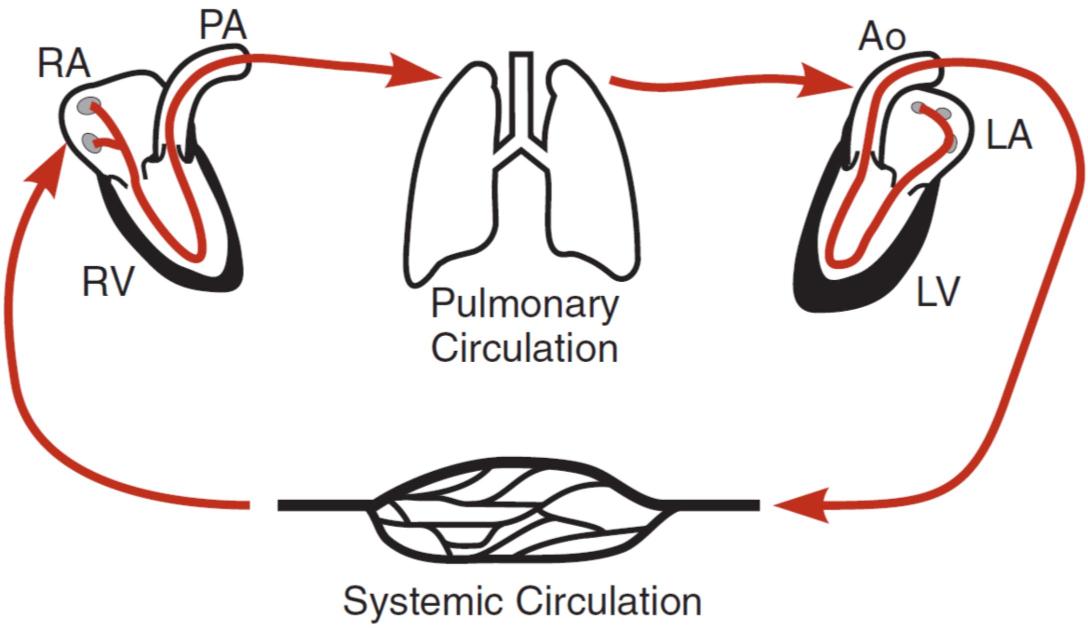

Figure 2.1: Schematic representation of cardiovascular circulation. RA: right atrium; RV: right ventricle; LA: left atrium; LV: left ventricle; PA: pulmonary artery; Ao: aorta. Adapted from [1].

2.1.1 The heart

The heart is divided in four chambers: right and left ventricle and right and left atrium. The right atrium is connected to the superior and inferior vena cava, receiving venous blood, returning from the systemic circulation, through them. This atrium can accommodate the incoming blood at low pressure due to its ability to highly expand. From there, the blood flows across the tricuspid valve, into the right ventricle. The blood then flows out of the ventricle, into the pulmonary circulation, through the pulmonary artery, which is separated from the right ventricle by the semilunar pulmonary valve. Returning from the lungs, the arterial blood enters the left atrium of the heart via the four pulmonary veins and then flows across the mitral valve, moving into the left ventricle. Due to its thick muscular wall, the left ventricle can achieve high pressures while contracting, causing the blood to cross the aortic valve and, consequently, be ejected into the aorta, once again initiating the systemic circulation [1].

2.1.2 Blood vessels

Allowing the blood to flow from and into the heart, the blood vessels are crucial in blood circulation. Due to their ability to constrict and dilate, blood vessels are essential in the regulation of arterial blood pressure, blood flow within organs, capillary blood pressure, and distribution of blood volume within the body, providing the required blood flow in all body tissues. Autonomic nerves, biochemical and metabolic signals from outside the blood vessel, and vasoactive substances released by endothelial cells that encircle the blood vessels are responsible for the adjustment of vascular diameters, through the activation of the vascular wall's smooth muscle. The substances produced by the cells in the endothelial lining also influence homeostasis and inflammatory responses [1].

Blood vessels are divided into five classes: arteries, arterioles, capillaries, venules, and veins. Arteries are characterized by their strong vascular walls, allowing blood under high pressure to flow at high velocity, and be delivered to the tissues. The arterioles, small branches at the end of the arteries, are also provided with strong muscular walls that allow these vessels to regulate the blood flow, by contracting completely or dilating, according to the needs of the tissues. Unlike the previous, the capillaries have very thin and porous walls that allow the exchange of fluids, nutrients, hormones, electrolytes and other small molecular substances between the blood and the interstitial fluid. From the capillaries, the blood is collected by the venules that progressively merge into veins. In turn, the veins not only conduct the blood from the venules to the heart but also serve as a reservoir for extra blood. Due to the low pressure present in the venous system, these vessels' walls are thin but still muscular enough to contract or expand in order to accommodate different amounts of blood that are not being needed in the circulation [14].

2.2 Blood pressure

Blood pressure represents a measure of the force it takes the heart to pump blood throughout the body. It is usually measured in units of millimeters of mercury (mmHg) and there are two types of blood pressure: Systolic Blood Pressure (SBP), that represents the peak aortic pressure measured when the blood leaves the heart, and Diastolic Blood Pressure (DBP), that is the measure of the minimal arterial pressure found right before the blood is ejected from the left ventricle into the aorta. Directly dependent on the two types of blood pressure mentioned is the pulse pressure, which represents the force the heart produces in each contraction and is measured as the difference between the systolic and diastolic pressures. Thus, any change in systolic or diastolic pressure directly affects pulse pressure [1].

Despite the clinical importance of the aforementioned pressure measurements, the pressure value that primary represents the blood flow driven to the organs is the average pressure over time, also called Mean Arterial Pressure (MAP). When at regular resting heart rates, it is possible to determinate the MAP value using the systolic (P_{sys}) and diastolic (P_{dias}) blood pressure values:

$$MAP \cong P_{dias} + \frac{1}{3}(P_{sys} - P_{dias}) \tag{2.1}$$

At higher rates, as blood is pumped into the systemic circulation and the pressure within the arterial vessels rises due to vascular resistance, the MAP is directly estimated by the Cardiac Output (CO), Systemic Vascular Resistance (SVR), and Central Venous Pressure (CVP):

$$MAP = (CO \cdot SVR) + CVP \tag{2.2}$$

The SVR, also referred as Total Peripheral Resistance (TPR), is the measure of the force that needs to be applied by the systemic vasculature on circulating blood. The CVP is the measure of BP in the vena cavas, right before entering the right atrium. At last, the CO is defined by the product of the Heart Rate (HR) by the Stroke Volume (SV) [1]:

$$CO = (HR \cdot SV) \tag{2.3}$$

Some of these concepts will be discussed in the next section.

2.2.1 Physiological factors

Some physiological mechanisms are known to influence blood pressure.

Stroke volume

Stroke volume refers to the volume of blood ejected in each contraction and is directly influenced by the energy of contraction of the ventricles and the pressure measured in the aortic and pulmonary arteries, two inversely related factors. When arterial pressure is elevated, stroke volume tends to be reduced, since the ejection can only happen after ventricular pressure surpasses aortic pressure. Under these conditions, most of the contractile energy is used to increase ventricular pressure, and less energy is available for ejecting blood. On the other hand, increased contraction energy, either by a greater sympathetic activity and circulating adrenaline or stretching of the myocytes due to an increased pressure at the end of the diastole (Frank-Starling mechanism), increases stroke volume [15].

BP is directly affected by the stroke volume, since a greater stroke volume increases the amount of blood that is ejected from the heart and needs to be accommodated in the arteries, resulting in a greater fluctuation in pressure during systole and diastole [14].

Heart rate

Heart rate represents the number of contractions per minute. Sympathetic or parasympathetic activation can increase or decrease heart rate, respectively. The former mediates cardiovascular functions through the release of noradrenaline, being responsible for an increase in cardiac conduction velocity and contractility, leading to an increase in heart rate. On the contrary, the latter can slow heart rate through the release of acetylcholine, a neurotransmitter capable of reducing contractility, opposing sympathetic activity in the heart [16, 17]. Even though elevated heart rate is usually associated with elevated BP, heart rate variability does not seem to have any direct relation with BP variation. For example, an increase in heart rate does not necessarily increase BP, since blood vessels have the capacity to dilate and accommodate the incoming blood without a excessive raise in BP [18, 19].

Cardiac output

Measured in L/min, the cardiac output represents the volume of blood that exits the heart through the ventricles. As described by the equation 2.3, increases either on heart rate or stroke volume can lead to an increase of the cardiac output and, consequently, elevate BP. To maintain normal BP levels, cardiac output and systemic vascular resistance must be stabilized. The onset of hypertension is normally associated to a raised cardiac output which results from an increased sympathetic activity, while the peripheral resistance is at a normal level [20, 21].

Systemic vascular resistance

Commonly known as peripheral vascular resistance, it refers to the vessels' compliance, that is the ability to expand and accommodate bigger volumes of content. The greater the compliance of a vessel, the more efficiently it can expand to accommodate a bigger volume of blood flow without increasing the blood pressure. Usually, veins are more compliant than arteries. Some vascular diseases, such as arteriosclerosis, can cause stiffening of arteries, reducing their compliance and increasing the resistance to blood flow. Hence, it results in greater turbulence, greater pressure within the vessel and reduced blood flow, increasing the work of the heart. Peripheral resistance is determined by evaluating the work of the walls of small arterioles, containing smooth muscle cells. A continued contraction of the smooth muscles can lead to a permanent increase in peripheral resistance by thickening of the vessel walls, what is thought to be mediated by angiotensin. Hypertensive patients normally present a regular cardiac output but an elevated peripheral resistance, believed to be triggered as a response mechanism to the increased cardiac output [20, 21].

Volume of circulating blood

Amount of blood flowing in the circulation. Specialized baroreceptors, located in the walls of the atria, can sense the increase of returning blood in the case of increased blood pressure, causing a stimulation of sympathetic activity and inhibition of parasympathetic activity to increase heart rate. The opposite is also verified [20].

Blood viscosity

Measure of the blood's thickness. It is directly influenced by the amount of proteins and other constituents present in the blood. An increase in blood viscosity causes resistance to flow, having a huge effect on BP [20].

Elasticity of vessel walls

The capacity of the vessel to resume its normal shape after stretching or compressing. Due to the abundance of elastic fibers in their composition, most larger vessels are typically elastic, allowing them to expand as blood flows through them and to constrict after. The more rigid the vessels, the lesser its compliance and greater the resistance to blood flow, which would require a lot more work of the heart to increase stroke volume and maintain an adequate pressure and flow, increasing BP. With low elasticity and increased pressure, vessel walls would have to become thicker [20].

Renal function and the Renin-Angiotensin-Aldosterone System

Renin is produced when the arteries that carry blood to the kidneys are narrowed and there is an insufficient blood flow in the area, when there is a reduced sodium intake, or simply by stimulation from sympathetic activity. Renin is a hormone that controls the production of angiotensin II, a vasconstrictor that affects the width of the arteries and stimulates production of aldosterone, that in turn regulates sodium and water retention. This is called the Renin-Angiotensin-Aldosterone System (RAAS), an important mechanism for regulation of the kidney's pressure-volume homeostasis, being necessary in states of low extracellular fluid volume and suppressed when there is an increase in extracellular fluid. RAAS is a system that widely controls BP by regulating many factors that influence it, such as sodium retention, reabsorption and excretion, salt sensitivity (a condition characterized by an increase of at least 10 mmHg in SBP, after a few hours of ingesting a certain amount of sodium), vasoconstriction, fluid volume and endothelial dysfunction (which causes diminish in the production of the vasodilator molecule nitric oxide and, consequently, increases BP) [14, 21–23].

Sympathetic nervous system

Sympathetic Nervous System (SNS) can induce both vascular dilatation and constriction and is an important system in BP regulation, even in cases when the changes in BP are only momentary (e.g., during physical activity). When the carotid artery is stretched due to the effect of high BP, the baroreceptors located in the carotid sinus induce a reduction in sympathetic activity in order to reduce BP. Hypertensive patients also manifest a decrease in parasympathetic activity. Even though there is no clear evidence that points to SNS hormones as a cause of hypertension, sympathetic overactivity is associated with endothelial dysfunction, vasoconstriction, and an increase in arterial stiffness, salt sensitivity and sodium reabsorption, and may be related to the onset and perpetuation of hypertension [21, 23].

2.2.2 Non-physiological factors

There are also many lifestyle factors that play a role in the alteration of one's blood pressure. Lack of physical activity, overweight and obesity, excessive salt and sodium consumption, alcoholism, smoking, and stress, stand out as modifiable life habits that negatively influence blood pressure.

Physical Exercise

Studies indicate that physical exercise helps control BP and is inversely related to the incidence of cardiovascular diseases [24–27]. Physical exercise triggers several physiological responses in the body to maintain the cellular homeostasis in the presence of increased metabolic needs. These physiological effects can be divided into three types: immediate acute, when the effects occur during and immediately after the physical exercise periods (like the rise in cardiac frequency, pulmonary ventilation, sudoresis); late acute, when the effects occur along up to 72h after the session (such as reduction on BP levels – especially in hypertensive people -, and increase of plasma volume, endothelial function and insulin sensitivity in skeletal muscles); and chronic effects, when the effects result from regular physical exercise (e.g., muscular hypertrophy, physiological left ventricular hypertrophy, promotion of angiogenesis, rest relative bradycardia and increased maximal oxygen intake). Briefly, the human body is subject to many cardiovascular and respiratory changes during an exercise period with the sense to respond to the demands of the active muscles. With the ongoing repetition of these changes, the performance of the organism is improved. The cardiovascular system is greatly influenced by the morpho functional effects of physical exercise, since it promotes increases on cardiac debt and redistribution of the blood flow, both as a response to the need to have a greater circulatory perfusion into active muscles. The resistance to blood flow also drops considerably with the start of exercise sessions, SBP increases directly proportionally to the raise on the cardiac debt and DBP varies with the capillary density of active muscles, influenced by the efficiency of their vasodilator mechanisms. Due to the various physiological and metabolic processes involved in the optimization of blood and oxygen distribution through active tissues during exercise, it is concluded that the drop in blood pressure after physical exercise is associated with those hemodynamic, humoral, and neural processes [28].

Sodium consumption

Sodium is directly related to changes in blood volume and BP, since high serum sodium concentrations promote water retention. A notable increase in sodium consumption provokes a series of hemodynamic reactions in an attempt to compensate for changes in BP levels. The most noticeable changes include a reduction in peripheral and renal vascular resistances and an increased segregation, by the endothelium, of nitric oxide, a vasodilator. The major risk factor of this process is associated with endothelial dysfunction. Chronic salt ingestion can lead to endothelial dysfunction, which in turn can result in the development of salt sensitivity. Salt sensitive individuals are subsequently prone to develop hypertension [23].

Weight

Overweight and obese patients are prone to arterial hypertension and present a higher resistance to treatment. Conversely to the majority of hypertension cases, where the arterial hypertension is often characterized by an increase in peripheral vascular resistance, obesity-related hypertension tends to be driven by an increased cardiac output, that happens to be mediated through expansion of plasma volume and sodium retention. Even though the blood pressure reduction that results from weight loss may be caused by the accompanying dietary manipulation and lifestyle changes, there is a direct relation between the two. Weight reduction towards a healthy body weight is proven to decrease BP, and, although the optimal interval is unclear, maintaining a healthy Body Mass Index (BMI) and waist circumference is beneficial for the reduction of BP, as well as for the prevention of hypertension in non-hypertensive individuals [6, 29, 30].

Alcohol consumption

Excessive alcohol consumption is strongly related to the increase of BP and the increased risk of hypertension and other cardiovascular diseases. Binge drinking has a great pressor effect since alcohol can affect blood pressure through multiple mechanisms. Acute alcohol consumption causes an increase of plasma renin activity, which in turn influences the RAAS. Increased plasma renin leads to an augmented production of angiotensin II, which functions as a vasoconstrictor and stimulates the adrenal gland to secrete aldosterone and vasopressin, enhancing sodium and water retention. All this results in an elevated BP, caused by an increase in peripheral resistance and blood volume. Alcohol reduction is therefore beneficial for cardiovascular health [6, 31].

Smoking

Smokers, both hypertensive and not, are prone to present higher daily BP values that non-smokers. Nicotine, the principal psychoactive constituent present in tobacco products, works as an adrenergic agonist in the organism, promoting the release of catecholamines into the circulation. High circulating catecholamine levels cause positive inotropic effects similar to those caused by the sympathetic nervous system, leading to stronger muscular contractions in the heart, that increase cardiac output and raises BP. Smoking is therefore one of the main contributions to the burden of cardiovascular disease and smoking cessation one of the most effective lifestyle changes to make in order to lower BP [1, 6, 32].

Stress

Exposition to stressful factors activates the SNS, which in turn may lead to elevations in cardiac output, heart rate and peripheral vascular resistance. All this provokes acute elevations in BP and a prolonged exposition to stressful factors may, in the long term, lead to chronic elevations in BP [33].

2.3 Hypertension

Hypertension is referred to as the incidence of a persistently high blood pressure in the arteries and it is defined as seated in-clinic measurements greater than or equal to 140 mmHg for SBP and/or greater than or equal to 90 mmHg for DBP. Nonetheless, it is recommended that BP and hypertension grades be classified in seven distinct levels, based on SBP and DBP, even for people that do not suffer from hypertension (table 2.1). BP categories are defined according to patient's seated office BP measurements and by the highest BP level, either systolic or diastolic. The classification is valid for people aged from 16 years, regardless of sex [6].

Table 2.1: Classification of office BP and hypertension grades. Adapted from *"2018 ESC/ESH Guidelines for the management of arterial hypertension"* [6]

Category	SBP (mmHg)	DBP (mmHg)
Optimal	< 120	< 80
Normal	120 - 129	80 - 84
High normal	130 - 139	85 - 89
Grade 1 hypertension	140 - 159	90 - 99
Grade 2 hypertension	160 - 179	100 - 109
Grade 3 hypertension	≥ 180	≥ 110
Isolated systolic hypertension	≥ 140	< 90

Depending on what caused the disease, hypertension can be divided into primary and secondary hypertension. Primary hypertension is the most common type of the disease (an estimated 90-95% of hypertensive patients suffer from this condition) and for most people it is a chronic condition with no treatment. Nonetheless, it needs to be managed, since it can cause damage to organs (e.g., heart and kidneys) and blood vessels. Its causes are unknown, and the diagnostic is made by exclusion, i.e., its diagnostic is made only after the known causes of secondary hypertension are excluded. However, there are theories that suggest that there is a relation between this type of hypertension and the kidneys' inability to efficiently manage sodium, since an increased sodium retention could be a sufficient factor for a higher blood volume and, consequently, a higher cardiac output that may eventually lead to an increase in systemic vascular resistance. Primary hypertension is also associated with age, race, diabetes and obesity, abnormal response to stress, heredity, and socioeconomic status. For secondary hypertension, it has a lot lesser prevalence (5-10% of cases) but it presents identifiable and treatable causes, as renal disease, primary hyperaldosteronism, sleep apnea, and some other diseases [1].

2.4 Time series

Time series are described as a set of observations, *x(t)*, of a variable, each sample recorded at a specific time t. There are two types of time series: continuous time series and discrete time series. The former is obtained when the measurements are made continuously through some time interval and are normally used to observe continuous variables, such as physiological signs (e.g., electrocardiogram). The latter are verified when the observations are taken at a discrete set of equally distant points in time. Discrete time series can be derived from continuous time series by sampling, in which there is an extraction of a given set of points at equal time intervals (for example, hourly interval measurements), or by aggregation over a period of time (e.g., total maximum temperature over a month). Currently, the devices used to record time series are mostly digital, producing signals in discrete time. If not, derivation of continuous time series is the usual approach for its analysis, since information loss can be minimal when considering the right sampling interval or aggregation period, and the complexity of the problem is significantly reduced [34, 35].

Time series prediction can be stated as a function $f : \mathbb{R}^N \rightarrow \mathbb{R}$ that uses the N time steps back from time t to obtain an approximation of the value of x at time t + p, with p being the prediction horizon, the number of time steps forward from t [36]:

$$x(t+p) = f(x(t), x(t-1), ..., x(t-N+1)) \tag{2.4}$$

Generally, this prediction aims to minimize the error between the estimated values, $\hat{x}(t+i)$, and the real values, $x(t+i)$, with $i = [1, 2, ..., p]$, being the most accurate prediction obtained when the result of the sum of squared errors (SSE), described in equation 2.5, is minimum [2, 37].

$$SSE = \Sigma_{i=1}^{p}(x(t+i) - \hat{x}(t+i))^2 \tag{2.5}$$

The rest of the present section focuses on the introduction, with special emphasis on their architectures and methodologies, of some models collected from the literature, not only the ones that will be developed in this study and those with the best overall performances, but also a few that are commonly used for baseline comparisons or that, although not yet widely implemented, present promising developments in this field. Linear regression will be considered as a simple linear method for baseline comparison with the remaining non-linear models, and even though there are many other linear and nonlinear models that are typically applied in time series prediction, their performances were not considered relevant to the present study, according to literature review.

2.4.1 Linear regression models

Linear algorithms assume that the relation between input features and target outputs is linear. Autoregressive models are a simple example of these type of algorithms and are normally used as a baseline for time series prediction.

Autoregressive models

Autoregressive models can predict future values of a time series through a regression of past values of that same time series. These models use a linear combination of input variables (x_{t-p}) to obtain the series value at time t:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... + \alpha_p x_{t-p} + \epsilon_t, \tag{2.6}$$

where x_t is denominated an autoregressive process of order p; α_p are the model's parameters (weights), estimated by minimizing SSE (2.5); and ϵ_t a random error (bias) [37, 38].

2.4.2 Nonlinear regression models

Unlike linear algorithms, nonlinear algorithms are based in the assumption that features and targets are related in a more complex and non-linear way. For example, deep learning models are nonlinear models commonly used in this context. This type of models can learn complex data representations from the data alone, without the need for human intervention in feature engineering, and provide accurate time series predictions. This segment presents a description of some deep learning architectures and other nonlinear models applied in time series prediction.

Jump Neural Network (JNN)

JNN [39, 40] is an alternative to the traditional deep feed-forward network. Deep feed-forward neural networks comprise an input layer, an output layer and multiple hidden layers, each containing a determined number of neurons. The input layer is connected to the first hidden layer and each hidden layer is subsequently connected to the next layer, with the last hidden layer connected to the output layer. Every neuron in a hidden layer is connected to each and every neuron in the next hidden layer. The same happens with input and output nodes and the neurons of the first and last hidden layers, respectively. Both the number of neurons and hidden layers are hyperparameters that affect the network's architecture and must be selected beforehand. In prediction problems, the input data is composed by values measured at past time instants. The described architecture is shown in figure 2.2.

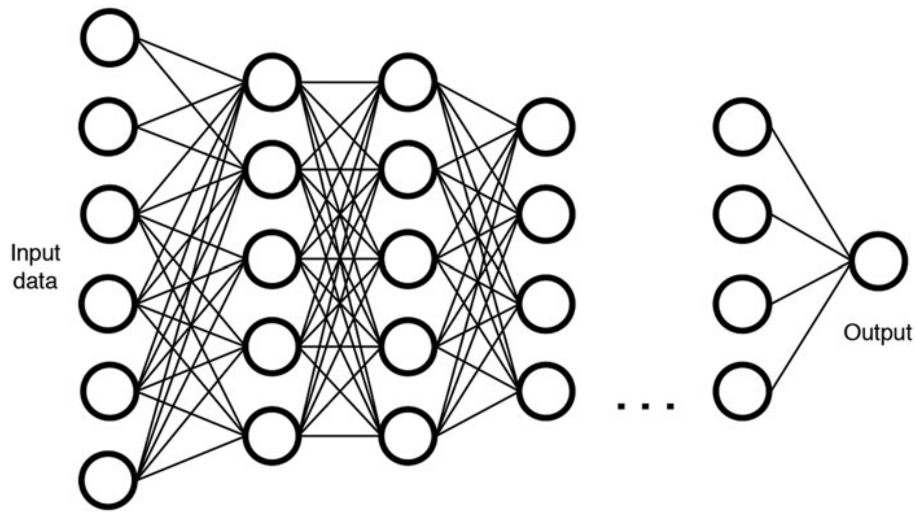

Figure 2.2: Example architecture of a deep feed-forward network. Adapted from [2]

During the training phase of the network, a gradient descent optimization algorithm is used to compute weights by minimizing the value of a loss function. These weights are then responsible for establishing relationships between neurons of two consecutive layers, which allows the computation of the output neurons values by a feed-forward operation:

$$a^l = g(W_a^l a^{l-1} + b_a^l). \tag{2.7}$$

This equation is true for every l-th layer, with a^l being the vector containing the values of the neurons in that layer (activation values), W_a^l and b_a^l the weight and bias, and g the activation function of the l-th layer. As seen, the activation values (a^l) of each layer are computed using the activation values of the previous layer (a^{l-1}) as input [2].

In the specific case of JNN, besides affecting the output through the hidden layer (n_k), inputs (x_i) are also directly linearly linked to outputs (y), in what is called a jump connection [3]. The mathematics for this network is as presented:

$$n_k = w_{k,0} + \sum_{i=1}^{i^*} w_{k,i} x_i, \tag{2.8}$$

$$y = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k N_k + \sum_{i=1}^{i^*} \beta_i x_i, \tag{2.9}$$

where i^* and k^* represent the number of inputs and neurons, respectively, β the coefficients related to the linear relationship, and w_k and γ_k the coefficients of the nonlinear relationship, with $w_{k,0}$ and γ_0 representing bias. At last, N_k is the activation function in the hidden layer [3]. For example, for a sigmoid activation function:

$$N_k = \frac{1}{1 + e^{-n_k}}. \tag{2.10}$$

A representation of this network's architecture, for three input values from past instants ($x1$, $x2$ and $x3$) and one hidden layer with two neurons, is shown in figure 2.3. y is the predicted output.

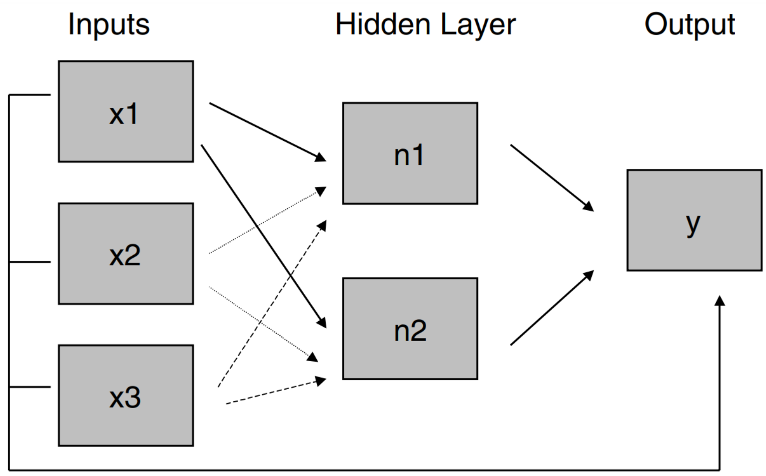

Figure 2.3: JNN network. Adapted from [3].

Radial Basis Function (RBF) networks

A RBF network presents two layers, and its architecture is similar to the one of a two-layer feed-forward network. Each unit in the hidden layer has a prototype vector (centroid) that is compared to the input vector. The output of each node is represented by a nonlinear function, the basis function, of the distance between the two vectors. Multiple types of basis functions exist, being Gaussian function one of the most used. The network's output functions can be both linear or non-linear, being the second one more computationally demanding and complex, as the output weights need to be optimised. The next equation describes the network output for linear output nodes:

$$y_k(x) = \sum_{j=1}^{M} w_{kj}\phi_j(||x - u_j||) + b_k, \tag{2.11}$$

where $M \in \mathbb{N}$ is the number of nodes in the hidden layer, $x = [x_1, x_2, ..., x_d]$ is the input vector, u_j is the prototype vector that defines the centre of the basis function ϕ_j, w_{kj} are the weights of the final layer and b_k the bias [4, 41]. The typical architecture of a RBF network is shown in figure 2.4.

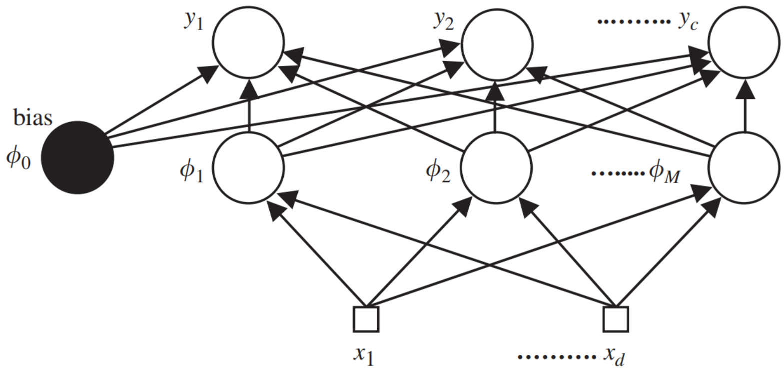

Figure 2.4: RBF network configuration. Adapted from [4].

Long Short-Term Memory (LSTM)

LSTM networks are a type of Recurrent Neural Network (RNN). Unlike feed-forward neural networks, where data is routed only from input to output, RNN connections contain a feedback loop whereby the current input is combined with the results from the hidden layer in the previous time step to form the new hidden layer input [42]. The architecture of a RNN is presented in figure 2.5.

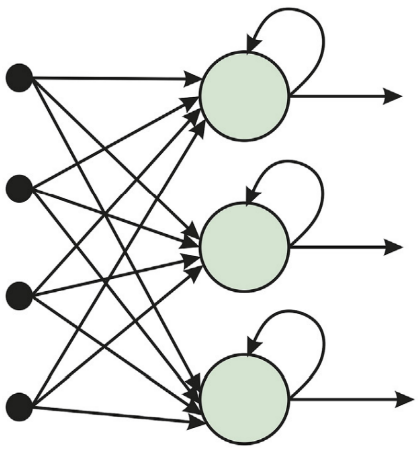

Figure 2.5: Example architecture of a RNN. Adapted from [5].

Due to the replacement of the activation functions by LSTM units, LSTM networks are RNNs optimized for long-term prediction on account of their capacity to process long time series data, regardless of the time span between significant events [42]. LSTM units incorporate memory cells, each consistent of a forget gate (Γ^f), an update gate (Γ^u), and an output gate (Γ^o). The gates alter the cell's memory state (c_t) by perceiving information as being useful or not, allowing the discarding of inconsistent information from the cell and solving the vanishing gradient problem (decrease of the gradient caused by the increase of the number of layers) affecting RNNs, allowing memory blocks to process and predict events with longer horizons. As the names indicate, Γ^f decides either information is saved or discarded, Γ^u decides if new information ($\tilde{c}_t$) will be used to update c_t, and Γ^o decides the value parsed as output to be used as input in the next hidden unit [2, 43, 44]. At last, the functioning of the LSTM units is defined by the input and output activation functions, usually the *tanh* and sigmoid (σ) functions, that compute the new information ($\tilde{c}_t$) and all of the mentioned gate values, respectively, using the input information coming from the previous hidden unit (a_{t-1}), along with the information from the current input (x_t):

$$\tilde{c}_t = tanh(W_c[a_{t-1}, x_t] + b_c), \tag{2.12}$$

$$\Gamma^u = \sigma(W_u[a_{t-1}, x_t] + b_u), \tag{2.13}$$

$$\Gamma^f = \sigma(W_f[a_{t-1}, x_t] + b_f), \tag{2.14}$$

$$\Gamma^o = \sigma(W_o[a_{t-1}, x_t] + b_o), \tag{2.15}$$

$$c_t = \Gamma^u \times \tilde{c}_t + \Gamma^f \times c_{t-1}, \tag{2.16}$$

$$a_t = \Gamma^o \times tanh(c_t), \tag{2.17}$$

where W_k and b_k, with $k = [u, f, o]$, represent the weights and bias of the respective Γ_k gate, W_c and b_c the weights and bias associated with the new memory state contender $\tilde{c}_t$, and c_{t-1} the previous memory state. a_t represents the values of hidden layer's neurons in state t [2]. A schematic representation of a LSTM hidden unit is presented in figure 2.6.

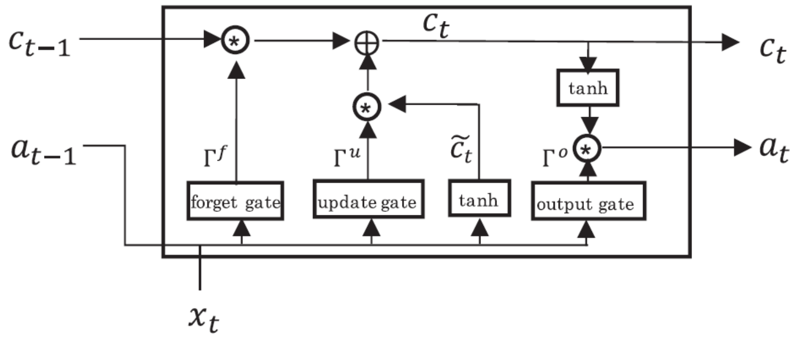

Figure 2.6: LSTM hidden unit. Adapted from [2].

Gated Recurrent Unit (GRU)

Another type of RNN, GRU is very similar to LSTM, but has a less complex structure, in comparison. Besides lacking an output gate, this type of architecture presents an update (Γ^u) and a relevance or reset gate (Γ^r) that control the information that is passed as output [2, 45].

Γ^u is responsible for deciding if the memory state (c_t) will be updated by the new contender to memory state ($\tilde{c}_t$) and Γ^r defines how much of existing memory is kept, that is, the relevance of the previous memory state (c_{t-1}) for the computation of the next contender to memory state ($\tilde{c}_t$). These processes are defined by the following equations:

$$\Gamma^u = \sigma(W_u[c_{t-1}, x_t] + b_u), \tag{2.18}$$

$$\Gamma^r = \sigma(W_r[c_{t-1}, x_t] + b_r), \tag{2.19}$$

$$\tilde{c}_t = tanh(W_c[\Gamma^r \times c_{t-1}, x_t] + b_c), \tag{2.20}$$

$$c_t = \Gamma^u \times \tilde{c}_t + (1 - \Gamma^u) \times c_{t-1}, \tag{2.21}$$

$$a_t = c_t, \tag{2.22}$$

where W_u, W_r, W_c, and b_u, b_r, b_c are the weights and bias of the Γ^u, Γ^r, and $\tilde{c}_t$, respectively. As described for LSTM, x_t is the input and a_t the values of hidden layer's neurons in state t. A GRU cell is shown in figure 2.7.

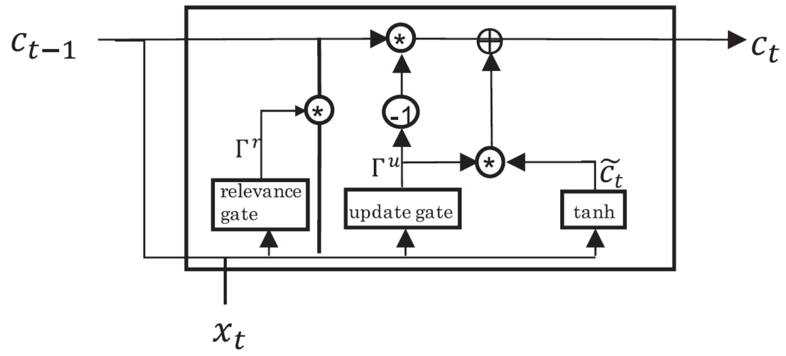

Figure 2.7: GRU hidden unit. Adapted from [2]

Case-Based Reasoning (CBR)

CBR models involve memory processes to integrate problem learning, understanding, and solving. This type of method is based on the assumption that similar problems have similar solutions, and its operation can be summarized in four procedures: retrieve, reuse, revise and retain. Basically, CBR searches for known cases that are similar to a new unknown case (retrieve) and uses their information to learn how to arrange a solution for the new case (reuse). For that, interpreting the new case at hand is crucial for the reasoner to find relevant cases that can be adapted to solve a problem. The model can form a novel procedure to deal with each new case. If the procedure has a successful execution and is found appropriate to solve the problem (revise), it can be indexed in the reasoner and retrieved when found useful (retain). Later on, if the same procedure is inaccurate in dealing with a new situation, it can be refined, resulting in an incremental learning process through feedback, improving the model's capacity to accurately deal with new problems [46, 47].

For simplicity of explanation, input data from a certain dataset will be referred as problem and its outputs as solution. Time series data, TS, from each subject is therefore organized in $(problem, solution)$ pairs named cases, (x_t, y_t), with x_t being the N observations before t, $x_t = [TS(t-N+1), ..., TS(t-1), TS(t)]$ and y_t the time series observations after t, up until a prediction horizon P, $y_t = [TS(t+1), TS(t+2), ..., TS(t+P)]$. For a new problem to which the solution is not known, the CBR model establishes a distance-based similarity search (usually using distance functions such as euclidean distance, Minkowski distance, etc.) to find similar problems in the database. A given number of the most similar cases is retrieved from the process and the values at each instant of the different solutions are processed (by averaging, weighted averaging, or other methods) to obtain the value of the respective instant in the new solution. For example, taking averaging as the adaptation method, the solution for the new case, $y_{new,t}$, will be obtained as follows:

$$y_{new,t} = \frac{\sum_{j=1}^{K} y_{j,t}}{K}$$

with K being the number of retrieved most similar cases, and $y_{j,t}$ the values of their known solutions at each instant $t = [1, 2, ..., P]$ in the prediction horizon. This makes CBR a personalized model that adapts to changes, since a new model is created for each new situation [47].

2.5 Summary

The cardiovascular system is composed by the heart and the vascular system, and ensures the necessary blood flow to the organs, facilitating exchange of gases, fluid, electrolytes, large molecules, and heat between cells and the and the surrounding environment [1].

An indicator of the proper functioning of this system is BP, the measure of force it takes the heart to pump blood throughout the body. BP can be greatly influenced by many physiological and non-physiological factors, each mediated by different mechanisms and having distinct impacts in BP.

A persistently high BP incidence, triggered by the action of adverse factors, can lead to hypertension, a chronic disease of great prejudice to the organism. Even though most cases have no treatment, the disease can be controlled by monitoring of BP levels, which allows timely action to reduce negative effects.

Regular measurement of BP can be described as a time series variable and, therefore, the early detection of discrepant BP values can be treated as a time series prediction problem. Such task can be accomplished through the implementation of certain machine learning models.

In this study, four of the mentioned models will be considered for development: LR, LSTM, JNN, and CBR. According to literature, these models have proven useful in predicting time series. LR was selected as the simple baseline linear model, LSTM the nonlinear model and JNN a nonlinear model with great influence from the linear component of the input signal. These three models allow a comparison of the influence of the variable's linear and nonlinear dynamics in the prediction. At last, CBR can be useful for assessment of prediction that is not based on the dynamics of the physiological signal.

3

State of the art

This chapter provides a brief explanation of the current state of the art on BP prediction, recommendation models and telemonitoring in general. Section 3.1 refers to telemonitoring strategies for home blood pressure management. Work related to the prediction module of the study is presented in section 3.2, including existent datasets (3.2.2), and prediction algorithms and strategies (3.2.1). Section 3.3 provides a state of the art of the recommendation module, that includes existing guidelines for hypertension management (3.3.1) and knowledge-based recommendation systems (3.3.2).

3.1 Telemonitoring and Home Blood Pressure

Telemonitoring consists in the use of information technologies to remotely monitor patients' conditions, interpret data, and make clinical decisions, transforming occasional office management in continuous management that extends into the patients' routine. With technological development, multiple telemonitoring modalities already exist but, as mentioned, the focus of this study is patient's self-management, which allows a frequent collection of clinical data that are easily measured by patients (e.g., weight and BP), and its transfer from the patients' home to the health care or telemonitoring center [48, 49].

Remote management has proved to be more effective than usual office care and it is mainly applied in the management of chronic diseases, like primary hypertension. It can reduce diseases' complications due to a better follow-up that allows early detection of worsening symptoms as well as early intervention. Accordingly, this setting is commonly advised to hypertensive patients with uncontrolled blood pressure, multiple comorbidities, and high risk of developing cardiovascular diseases, as well as patients living in isolated regions and older patients, since it ensures continuity of care. It also allows the detection of masked hypertension, that is defined by normal ranged office BP readings and elevated home or ambulatory BP measurements, and white coat hypertension, that happens when the opposite is true [50, 51].

The use of telehealth strategies to assist hypertensive patients in reduction and management of BP was first specifically recommended only in the 2017 American Heart Association (AHA)/American College of Cardiology (ACC) clinical practice guideline for high BP [52]. Until then, major guidelines only issued partial or incomplete recommendation of these techniques [50].

There are currently multiple accessible services of this sort, but none seems to be a possible reference to form a standard telemedicine healthcare model, since they are usually not interoperable or integrated into existing healthcare systems. However, several studies have noted that the employment of telemonitoring to the detriment of usual care approaches resulted in a more significant BP reduction and control, which is possible due to an improvement in patient engagement and knowledge on their condition, adherence to the care plan, and satisfaction with the intervention [50, 53].

Omboni et al. (2020) [53] aimed to review the clinical utility and barriers of the application of telemonitoring in hypertension management. For that matter, a systematic review and meta-analyse was conducted on fifteen studies on the efficacy of telehealth compared with usual care on hypertension management, based on type of intervention, type of subjects (division into subgroups differentiated by existent comorbidities), application setting (e.g. communities, hospitals, pharmacies, etc.), number of studies compared, number of subjects, evaluated results and other factors. It was concluded that the main benefits of the use of this type of strategy happened in services that established a remote monitoring of vital signs (e.g. BP) allied with adherence to the medication plan, education on healthy lifestyle habits and risk factors, and an asynchronous feedback (periodic interaction with the patient by the caregivers, instead of immediate). In terms of feedback and management, the best approach was assumed to be the automatic delivery of recommendations resultant from algorithms, mixed with the supervision of a clinical team (when needed). At last, the target population was concluded to be patients with suspected hypertension or white coat hypertension, hypertensive patients (especially those at high-risk, due to poor medication adherence, etc.), older adults, people living in deprived areas and hypertensive patients that present multimorbidity (e.g. diabetes mellitus, obesity).

Duan et al. (2017) [54] published a systematic review about the effectiveness of Home Blood Pressure (HBP) telemonitoring. The comparison of forty-six randomised controlled studies focused on the effectiveness of HBP telemonitoring conversely to usual care in reaching target BP and was centered on multiple outcomes: BP changes and normalisation, application of antihypertensive medication, and quality of life (mainly physical and mental health). The studies were analyzed based on duration of intervention and the existence of additional support, categorised into "counselling", when regular face-to-face or telecounseling communication is used for BP evaluation or exchange of information on disease management between patients and healthcare assistants; "education", which consists of providing information on self-management and non-drug therapies, such as healthy lifestyle changes; or "miscellaneous", an approach that involves both "counselling" and "education". That way, the authors were able to assess three main aspects of the approach: whether or not HBP telemonitoring was more beneficial in the reduction of office and ambulatory BP than usual care; if there was a difference in the efficacy of HBP telemonitoring in BP-lowering with or without additional support; and the effect of different intervention components (duration and additional support) on BP outcomes. For the first comparison, results showed an improve in office SBP and DBP of 3.99 mmHg and 1.99 mmHg, respectively, and a larger number of patients achieving BP normalisation when HBP telemonitoring is applied instead of usual care. For the second, telemonitoring with additional support presented a reduction of 2.44 mmHg and 1.12 mmHg more for SBP and DBP, respectively, than telemonitoring without additional support. For the latter, a lower BP was achieved when telemonitoring lasted between 6 to 12 months and counselling support was provided. The overall conclusion presented was that, accordingly to previous studies, HBP telemonitoring improves BP control and is more efficient in achieving target BP than conventional treatment. It also noted that the approach provides a significantly enhanced control of both ambulatory and office BP, when compared with usual care, and that the inclusion of the mentioned additional support in the HBP telemonitoring intervention benefits BP decrease.

In agreement with the above, other studies of the sort [50, 51, 55–58], have also reported similar results.

3.2 Prediction module

3.2.1 Prediction algorithms

In order to assist BP management, the prediction of BP is an increasingly important component for the early detection and prevention of cardiovascular disease. BP prediction facilitates the detection of high-risk BP values and other risk factors, allowing an early diagnosis of hypertension, even in healthy people, and alerting to the need to make preventive decisions in this regard [59]. Deep learning algorithms are the most commonly used techniques in existing researches on BP prediction.

Li et al. (2017) [42] presented a LSTM model, which allows the processing of sequences of inputs for an accurate BP prediction. The simultaneous use of a contextual layer in addition to the model allowed consideration of both measurements and contextual data for BP prediction. That layer was added as an extra hidden layer after the first hidden layer in the LSTM model, and takes as input both the features extracted from the measurement data in the first hidden layer and the contextual data. A linear regression of the outputs of the contextual hidden layer is used for prediction of SBP and DBP up to three months. A private dataset was used, containing measurement data concerning SBP, DBP and HR, and contextual data such as BMI, age, gender and geographic information. The output results were evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). For DBP and SBP, MAE ranged from around 3.23 to 4.24 mmHg and from 5.27 to 6.92 mmHg, and RMSE ranged from 4.27 to 5.54 mmHg and from 6.97 to 9.04 mmHg, respectively, which seem satisfactory results considering the normal range of DBP and SBP.

Koshimizu et al. (2020) [59] proposed the use of deep NNs to predict BP variability and mean value from time series data of HBP as well as medical data measured at a hospital, since previous studies have concluded that BP variability is an outstanding risk factor that leads to cardiovascular disease. The PREDICT dataset (described in 3.2.2) was used. Outputs of the prediction model consist of weekly BP mean values and variability (up to 4 weeks), and input data comprises patient's context data acquired once and time series data acquired over 56 days. Both LSTM and GRU models were developed, in which BP time series data were used as input to the deep neural networks and contextual data were input to a fully connected layer implemented in the models, in order to introduce a dependency between the time series and contextual data and obtain more accurate outputs. Also, a novel loss function was developed, to allow the simultaneous evaluation of both BP mean values and variability prediction errors and increase prediction accuracy. Favorable results were obtained, with variability evaluation using standard deviation ratio (SR) ranging between 0.67 and 0.70 and mean value evaluation using RMSE ranging between 5.04 and 6.65 mmHg.

Liu et al. (2018) [43] suggested an extension to the traditional LSTM approach, by incorporating a RBF for contextual information processing. The model adopted a multi-task training strategy, allowing the simultaneous prediction of BP and other related time series, processed in the same two hidden layers. The contextual information (age, BMI, etc.) is input to the RBF NN, that is then used as the transfer function of the model's hidden layer. This allows the recurrent structure to extract features from BP and BP related data. Both MAE and RMSE were the considered evaluation metrics. For short-term prediction, the study's model was compared to the results obtained by a traditional LSTM model and a LSTM model with a contextual layer (proposed in [42]) with the same data, outperforming both. Long-term prediction, in which the model was developed using 30-day continuous BP data to predict the next 3-day's BP, resulted in accurate sequential prediction values up to 3 days, with MAE ranging from 4.52 to 4.66 mmHg and from 9.68 to 9.97 mmHg and RMSE ranging from 6.21 to 6.40 mmHg and from 12.76 to 13.01 mmHg, for DBP and SBP, respectively.

Zecchin et al. (2014) [39] presented an algorithm for blood glucose concentration short-time prediction. Theoretically, a similar approach can be applied when dealing with BP time series. The novel model, named Jump Neural Network (JNN), consists of a deep feed-forward NN, that differs from the previously described architecture by having the input layer connected to both the first hidden layer and output layer. Since the considered blood glucose time series follows both linear and non-linear dynamics and glucose signals' inputs are able to influence future effects, the model's structure is suitable for data fitting and prediction, having the hidden and output neurons dealing with the non-linear and linear relationship between inputs and targets, respectively. In this particular case, the model also comprised information on ingested carbohydrates and obtained accurate predictions, similar to those obtained by models proposed in previous studies for the effect. The results were considered satisfactory for the study but their presentation is irrelevant in the present context.

Rocha et. al (2011) [60] developed neural network multi-models to predict the occurrence of acute hypotensive episodes in intensive care units. The initial phase of this approach relied on a CBR technique to find similarities between the input MAP signals and MAP templates that are available in the training dataset and represent evolution trends of MAP signals. Similarity search was based on correlation analysis. A set of neural multi-models, previously trained using the identified most similar templates, were then selected for prediction of future evolution of the current input MAP signal, allowing a direct determination of the hypotensive episode's occurrence. The described multi-model approach follows the regression representation presented in equation 2.4. These models were implemented using a Generalized Regression neural network (GRNN), which can be seen as a normalized RBF network, and is here used for MAP signals prediction. Both training and test datasets consisted of a sample of selected patient records from the MIMIC-II database. Training dataset consisted of 60 records with all available data, while the test sets were truncated at an instant T_0 and consisted of dataset A (10 records) and B (40 records). The forecast window was the hour immediately after the instant T_0. The proposed methodology was able to correctly predict 47 out of the 50 acute hypotensive events, presenting a global sensitivity (SE) of 94.74%, specificity (SP) of 93.55% and accuracy (AC) of 94.00%.

Prediction algorithms overview

To summarize, an overview of the mentioned prediction approaches, is presented in table 3.1.

Table 3.1: Overview of the mentioned prediction algorithms.

Study	Algorithm	Methods	Prediction horizon	Best results
Li et al. (2017) [42]	LSTM	An extra hidden layer after the first hidden layer combines features extracted from the measurement data and contextual data as input. Prediction of SBP and DBP based on linear regression of outputs from the extra layer.	Up to 3 months	SBP/DBP MAE: 5.27/3.23 mmHg RMSE: 6.97/4.27 mmHg
Koshimizu et a[59]	LSTM GRU	Combination of LSTM/GRU models for time series data with fully connected layer for contextual data, to predict weekly BP mean values and variability.	Up to 4 weeks	SR for BP variability: 0.67 RMSE for mean BP values: 5.04 mmHg
Liu et al. (2018) [43]	LSTM RBF	RBF network for contextual data processing used as the transfer function of the LSTM model's hidden layer. Multi-task training for prediction of BP and BP related time series.	3 days	SBP/DBP MAE: 9.68/4.52 mmHg RMSE: 12.76/6.21 mmHg
Zecchin et al. (2014) [39]	JNN	Prediction of blood glucose time series using a novel type of feed forward neural networks, the JNN.	30 min	-
Rocha et al. (2011) [60]	CBR GRNN	Similarity-search to compare MAP signals with existing MAP templates and prediction of acute hypotensive events using the correspondent trained neural network multi-models to forecast MAP signals.	1 hour	47 out of 50 events correctly predicted SE = 94.74% SP = 93.55% AC = 94.00%

3.2.2 Datasets

From the literature, several datasets are known to provide useful data for BP prediction. Although all the ones found are described in this section, the majority are private datasets, and only MyHeart and MIMIC-III are available to the team.

PREDICT

The PREDICT (Prediction of ICT-Home Blood Pressure Variability) dataset was created in the context of a trial conducted by Jichi Medical University in Japan, with the goal to assess the efficacy of the use of HBP, collected using information and communication technology, for prediction of cardiovascular events. Participating patients were asked to daily measure their BP at home for two or more years, using a device that was able to measure surrounding environment temperature, with the intent of forming an algorithm to identify thermosensitive hypertension. The measurements took place twice a day, in the morning and in the evening, with two measurements per occasion, to prevent possible measurement errors. Besides time series data, contextual data was also stored. The collected time series data consists of measurement date and time, number of measurements per day, SBP, DBP, HR, and contains other observations related to the measurement act, such as cuff fit, presence of irregular heartbeat, and body movements and temperature during measurement. For the patients context data, age, gender, height, weight, medical history (presence or absence of certain medical histories, such as myocardial infarction, stroke, heart failure, chronic arterial occlusion, etc.), present illness (presence or absence of certain illnesses, such as hypertension, kidney dysfunction, diabetes, sleep apnea syndrome, etc.), the use of certain antihypertensive prescription drugs, and drinking habits were collected [59, 61].

TEN-HMS

The TEN-HMS (The Trans-European Network – Home-Care Management System) study [8] comprised 426 patients over 18 years of age who were willing to comply with home telemonitoring. All patients included fulfilled certain conditions, such as an admission in a hospital within the previous six weeks on account of worsening heart failure lasting over 48h, chronic heart failure symptoms, left ventricular ejection fraction (percentage of blood pumped by the left ventricle with each contraction) under 40%, an antihypertensive drug prescription regimen containing over 40 mg/day of furosemide (a diuretic medicine used in the treatment of hypertension, heart failure and other conditions [62]) and other similar related criteria that increased heart failure risk. Social and demographic context data was gathered for this dataset. A baseline characterization of each patient was formed before initiating management containing their age, sex, primary cause of heart failure (e.g., hypertension, coronary disease, alcohol-related, etc.), existent comorbidities (e.g., hypertension, stroke, diabetes, etc.), a reference basis in terms of weight, BMI, SBP, and DBP, measurement values from a blood sample (such as hemoglobin, serum sodium and creatinine), among others. The patients were then instructed to perform a twice daily set of measurements, before breakfast and before their evening meal and medication, of their weight, blood pressure, and heart rate. The objective of this study was to assess, through a follow-up of 450 days, whether home telemonitoring strategies could improve the outcomes of heart failure patients at high risk of death or hospitalization, comparatively to usual care and nurse telephone support (similar to usual care, except patients were contacted monthly by a specialist nurse to evaluate their symptoms and medication) strategies [8]. Even though this study is not hypertension management oriented, the collected data is relevant in that context.

MIMIC-III

MIMIC-III (Medical Information Mart for Intensive Care) is a freely-accessible database containing a large amount of data from 2001 to 2012, relating to patients admitted to critical care units of a hospital of Harvard Medical School, in Boston. The database comprises deidentified health-related data including vital signs, medication, laboratory measurements, length of stay, survival data, imaging reports, some observation and notes from healthcare providers, among other information. The time-stamped physiological data available in the MIMIC-III consists of approximately hourly measured nurse-verified vital signs (e.g., blood pressure, heart and respiratory rate). It is a widely used internationally database, not only on academic research but also on an industrial level, since it is the only freely-available critical care database with a span of over a decade, that provides detailed information for each individual patient. Other previous and updated versions of the dataset are available, like MIMIC-II and MIMIC-IV, respectively [63, 64].

MyHeart

The MyHeart research project [65] was created with aim to establish preventive measures for cardiac disease using wearable solutions that allow premature diagnosis and timely intervention, in a continuous home telemonitoring setting. The project involved the research and development of new wearable technologies (such as sensors) and algorithms used in monitoring of various vital signs (e.g. blood pressure, heart rate, respiratory rate, and weight), expecting the early detection of heart failure decompensation symptoms. The database comprises 1 year worth of home measured data from 148 chronic heart failure patients. Patients were instructed to perform daily measurements of vital signs at rest, that always included a one-lead electrocardiogram signal [66].

IDHOCO

The IDHOCO (International Database of HOme blood pressure in relation to Cardiovascular Outcome) database [67] includes data from 6753 patients from 5 cohort studies on home BP monitoring. The purpose of the study was to determine the diagnostic of the patient outcome, based on self-measurement of HBP, research on the predictive value of factors such as morning and evening BP, white-coat and masked hypertension and BP and HR variability, and determine an optimal time schedule for HBP measurements to accurately assess the risk for cardiovascular events. Baseline characteristics of all patients at enrollment are included, namely age, sex, BMI, height and weight, number of home and office SBP and DBP measurements, home and office HR, serum total cholesterol, smoking habits, antihypertensive drug treatment, and the presence of diabetes mellitus, history of cardiovascular disease and hypertension. The registered office BP represents the average of two consecutive readings obtained in the sitting or supine position, after the patient had been resting for two or more minutes. HBP measurements were obtained by the participants at their home, while in the sitting position, after a two to five minutes rest and value of this variable on the database represents the mean of all available measurements [67–69].

CPCSSN

The CPCSSN (Canadian Primary Care Sentinel Surveillance Network) project [70] collected and validated data from electronic medical records across Canada, concerning multiple conditions, such as hypertension and diabetes. The dataset contains both provider and patient information. Patient data comprises information on demographics, health condition, medication, risk factors, allergies, physical examinations for SBP, DBP, height, weight, waist circumference, waist to hip ratio, BMI, and peak expiratory flow rate, and other variables related to medical visits and procedures and laboratory examinations [70, 71].

Datasets overview

To summarize, an overview of the mentioned datasets, containing relevant variables, number of patients, follow-up duration and accessibility is presented in table 3.2.

Table 3.2: Overview of the mentioned datasets containing useful information for BP prediction and hypertension management

Dataset	Relevant variables	Number of patients	Follow-up duration	Accessibility
PREDICT	SBP; DBP; HR; age; sex; height; weight; medical history; present comorbidities; antihypertensive prescription drugs use; drinking habits	Unknown (target sample size 1100)	At least two years	Private
TEN-HMS	SBP; DBP; HR; age; sex; weight; primary cause of heart failure; existing comorbidities	426	450 days	Private
MIMIC-III	BP; HR; respiratory rate; age; sex	Over 40 000	11 years	Public (upon request)
MyHeart	BP; HR; respiratory rate; weight	148	1 year	Private
IDHOCO	SBP; DBP; HR; age; sex; height; weight; BMI; smoking habits; serum total cholesterol; antihypertensive prescription drugs use; existing comorbidities	6753	Unknown	Private
CPCSSN	SBP; DBP; HR; age; sex; height; weight; BMI; waist circumference; waist to hip ratio peak expiratory flow rate	Unknown	Unknown	Private

3.3 Recommendation module

3.3.1 Guidelines for hypertension management

ESC/ESH 2018

The 2018 ESC/ESH guidelines [6] propose changes in lifestyle to treat and control hypertension, as well as to reduce the risk of cardiovascular disease. Several measures are recommended, including changes in diet, moderation of alcohol consumption, weight reduction, regular physical exercise, and smoking cessation.

In terms of diet, sodium is considered one of the principal factors with a pressor effect, being associated with the rise of SBP and risk of hypertension. Excessive sodium intake has been associated with an increase in SBP with age, leading to the prevalence of hypertension. Inversely, sodium restriction is proven to be efficient in lowering SBP and DBP. It is recommended a limitation of sodium consumption to 2g per day (approximately 5g salt). Reductions of around 1.75g in sodium consumption are associated with a decrease of up to 5.4 mmHg for SBP and 2.8 mmHg for DBP, in hypertensive patients. However, mainly due to poor dietary persistence, the BP-lowering benefits of sodium restriction tend to diminish with time. Other dietary habits are also known to be effective in prevention or delay of the onset of hypertension. Hypertensive patients are advised to follow a healthy balanced diet with increased consumption of vegetables, fruit, fish, whole grains, white meats, beans, nuts, and unsaturated fatty acids (like olive oil). Foods rich in potassium, calcium, magnesium, fiber, and protein are recommended, as well as low consumption of red meat, saturated fat, sugar, and fat dairy products. Dietary Approaches to Stop Hypertension (DASH) and Mediterranean diets are examples of recommended diets that follow the mentioned eating habits used in hypertension prevention and BP control.

Weight control is also recommended for maintaining a normal BP. Excessive body weight gain is strongly related with the prevalence of hypertension, and it has been proved that its reduction also decreases BP. An average weight loss of around 5 kg is empirically associated with a reduction of 4.4 and 3.6 mmHg on SBP and DBP, respectively. Weight reduction is an equitable goal for the overall population but it is an highly endorsed measure for overweight and obese hypertensive patients, since is particularly useful in stabilizing metabolic risk factors and can also improve the efficacy of antihypertensive drugs. Even though there is no consensus on the optimal BMI, maintaining a healthy body weight (BMI of around 20 to 25 kg/m2, or higher for older patients) and waist circumference (less than 94 or 80cm for men and women, respectively) is a beneficial measure in preventing hypertension and reducing BP. This process is associated with and highly dependant on other habits, such as dietary changes and regular physical exercise.

Physical exercise is responsible for a severe rise in BP (mainly SBP) during the activity and a subsequent temporary decline in BP below baseline. For general populations, moderate and high intensity training are more efficient in lowering BP than regular lower intensity physical activity, but the latter has been associated with a reduction of more than 15% in mortality. Aerobic, resistance and isometric training all have been proved to contribute to a decrease in resting SBP/DBP of 3.5/2.5, 1.8/3.2, and 10.9/6.2 mmHg, respectively. That said, hypertensive patients are advised to practice at least 30 min of moderate intensity aerobic exercises (e.g. walking, jogging, swimming) on at least 5 days weekly, with resistance exercise on a minimum of 2 days also showing great benefits. The impact of isometric exercises is less well known in this context.

Alcohol consumption is also considered to have a great pressor effect, causing an increase in BP and the prevalence of hypertension, binge drinking being the factor with the most influence in that regard. Even though studies have only shown a decrease of 1.2 and 0.7 mmHg in SBP and DBP, respectively, for an intervention group in a 6 month period, epidemiological studies verified that reductions in alcohol consumption are beneficial. Hypertensive patients are advised to avoid binge drinking and have alcohol-free days, while limiting alcohol consumption to 14 and 8 units weekly for men and women, respectively (1 unit corresponding to about 8g of pure alcohol).

Smoking is highly prejudicial for cardiovascular health, with both normotensive and hypertensive smokers presenting higher BP values than the average non-smokers.

Although no chronic effects of smoking were assessed in office BP, smoking cessation is appointed to be the most effective lifestyle measure for prevention of many cardiovascular diseases, being the referral to programs and supporting care regarding smoking cessation recommended during patient visits.

ACC/AHA 2017

The 2017 ACC/AHA [52] guidelines focus on overweight and obesity, sodium and potassium on diet, physical fitness and alcohol consumption. Overall, the same remarks are made on the present factors as in [6].

Relatively to weight, a notorious relationship between increased body weight and high BP and hypertension is stated. An almost linear relationship between BMI or waist-to-hip ratio and BP is also pointed. Existing studies on the matter even indicated that obesity may be responsible for up to 78% and 65% of hypertension cases in men and women, respectively.

Diet-wise, sodium and potassium intake are mentioned as having two opposite effects on BP. The first, as referred in [6], is positively associated with BP increase, while the latter appears to cut off the effect of sodium intake and many epidemiological studies suggested that a lower sodium-potassium ratio is liable to be associated with a reduced overall risk of cardiovascular disease.

Physical activity is once again noted as inversely related to BP level and hypertension. A cohort of men whose ages ranged from 20 to 90 years was also mentioned, since they were followed for a minimum of 3 and up to 28 years and the results proved that a higher physical fitness caused hypertension onset to be delayed, due to a minor rise in SBP over time.

Guidelines overview

Many guidelines from worldwide famous societies and hospitals specialized in hypertension exist [6, 52, 72–74]. Although most are in agreement with the above mentioned, some add other factors. To summarize, an overview of the cited clinical practice guidelines is presented in table 3.3.

Table 3.3: Overview of lifestyle modifications recommended for hypertension management by multiple clinical practice guidelines.

Healthy diet	Reduced sodium intake (less than 5g salt per day is recommended). Increased potassium intake, keeping a low sodium-potassium ratio. Patients are also advised to follow a balanced diet (such as DASH and Mediterranean diets) rich in calcium, magnesium, fiber, protein, vegetables, fruits, fish, nuts, and unsaturated fatty acids (e.g. olive oil), and low on red meats, sugar and fat dairy products.
Weight control	Maintenance of a healthy body weight (BMI between 20 and 25 kg/m2, slightly higher for older patients), and waist circumference (<94 cm and <80 cm in men and women, respectively). Waist-to-height ratio is also an alternative indicator (<0.5 is advised).
Physical activity	Aerobic exercise (walking, running, jogging, swimming, etc.) of moderate intensity for 30 min on at least 5 days a week or high intensity, involving short intervals of intense activity followed by recovery periods of lighter activity. Resistance training on 2 or 3 days a week is also beneficial.
Alcohol reduction	Avoid binge drinking and adopt alcohol-free days weekly. Alcohol consumption should be limited to 20g or 15g per day, and to 112g or 64g per week, for men and women, respectively.
Smoking cessation	It is recommended that patients cease smoking completely. Smoking causes an acute increase in BP that lasts for 15 min or more and is associated with the onset of hypertension. Referral to smoking cessation programs are advised.
Stress reduction	Even though chronic effects of stress in BP are not well determined, induced mindfulness is suggested to lower BP.
Other factors	Environmental factors like air pollution and cold temperatures have a negative influence on BP. By stimulating the activity of the sympathetic nervous system, sleep disorders also have an impact on the onset of hypertension.

3.3.2 Knowledge-based recommendation systems

Knowledge-based recommendation systems are useful tools for chronic patients monitoring, facilitating self-management of patients and improving adherence to treatments and guidelines. Even though there are other types of health recommendation systems, knowledge-based support systems make use of information previously established by experts, like the guidelines presented in 3.3.1, to give personalized advice to users according to the available data that meet users' preferences and needs, being the most useful health recommendation system method in the current context.

Vives-Boix et al. (2017) [75] presented a knowledge-based clinical decision support system for monitoring chronic patients, namely hypertensive patients. The study focused on four risk factors, common to multiple chronic diseases: diet, physical activity, alcohol consumption and smoking. The goal of the system was to analyse daily measurement of physiological parameters (such as blood pressure, age, height, BMI, HR and daily physical activity), in order to give personalized recommendations for management of the disease. System's rules involving lifestyle changes to decrease BP follow the guidelines described in section 3.3.1, specifically reduction of salt, alcohol and saturated fat consumption, weight reduction, increase of physical activity and consumption of fruits and vegetables, smoking cessation, and adoption of a Mediterranean diet. The knowledge-based system is centered on the provided information about the topic, in the form of IF-THEN rules. Then, an inference engine retrieves available rules and combines them with the users' data. At last, the results from the evaluation of users' data are presented as recommendations.

Wang et al. (2020) [76] aimed to develop and evaluate the performance of a knowledge-based recommendation system to provide personalized health advice for chronic disease patients. The system's goal was to recommended educational materials (obtained from various sources) depending on the patients' needs, assessed through their health data (e.g., age, gender, comorbidities, diet and physical activity levels, BP value, etc.), that was available in a dataset obtained from a telehealth system. In the process, both patient data and educational materials are inputted into an ontology system that converts them to vectors and the final recommendation is generated based on similarity between the two vectors. Given recommendations were subsequently evaluated manually by experts, that assigned whether they were appropriate or not. The developed system was implemented in a mobile health system and connected to patients' smartphones, returning a variable number of recommendations, depending on how many suited each input case. Evaluation of the top 1 recommendation reached a score up to 0.970 for macro precision and 0.628 for mean average precision.

Silveira et al. (2019) [77] developed and evaluated a knowledge-based mobile clinical decision support system for hypertension management. International clinical practice guidelines, namely some of the mentioned in 3.3.1, were used as the basis for the construction of recommendations. Users are required to manually input multiple data fields, including sex, date of birth, height, weight, BP, waist circumference, existent comorbidities and drug prescription, and other relevant laboratory test results. The inserted data is then evaluated by the system, namely for BMI and cardiovascular disease risk assessment, and recommendations on lifestyle interventions, such as physical activity, diet and medication dosages are presented to the user. The patient data evaluation and rules comparison process is not described in the article. For evaluation of the feasibility, usability and utility of the system, participants of the study were asked to fill some questionnaires or participate in interviews, to classify their general impression. The overall system classification was good and, even though 70% of the clinicians involved reported data input as time consuming, the majority of physicians found the system useful to optimize treatment and prevention of hypertension, as it was easily incorporated into daily routines, making it possible to be incorporated in a primary care setting.

Other systems of the sort were developed in multiple studies. As an example, [78] developed a knowledge-based recommendation system for diet and exercise advise as a prevention measure for chronic diseases and other illnesses, following similar approaches as the ones mentioned.

3.4 Summary

Based on the found studies mentioned in the state of the art, we can prove the efficacy of telemonitoring strategies like HBP telemonitoring in lowering BP, to the detriment of usual office care. Additionally, there is evidence that these approaches are associated with a greater incentive for patients to adopt healthier life habits and comply with treatments. This concludes that telehealth strategies can be useful supplements to interventions proved to reduce and control BP in hypertensive patients.

As for the prediction module, it is stated that multiple studies and datasets on the subject exist, proving that BP prediction, mainly using deep learning and other machine learning algorithms, such as LSTM, JNN and CBR models, is possible and represents a valid and useful strategy for hypertension management and prevention of cardiovascular diseases, that can be incorporated in a remote monitoring system.

Relatively to the recommendation module, existing international guidelines from different societies and association specialized in hypertension provide some basis information for the construction of rules for a knowledge-based recommendation system. The mentioned guidelines are in overall agreement with each other and identical information is used in some of the collected recommendation systems, which proves them useful in this context. At last, as the developed systems were found beneficial for hypertension treatment, management and prevention, and most of the systems are also employed in mobile health applications, the use of this type of systems in the telemonitoring approach discussed in the present study may be a viable approach.

4

Methodology

This chapter is divided in two parts. The first part covers the methodology conducted for the development of the prediction module, from data treatment to algorithms' architectures, training, hyperparameters tuning and performance assessment. The second part addresses the recommendation module, the establishment of the set of rules for the knowledge-based recommendation system and its construction.

Both developed modules are intended to be implemented, jointly with the modules from other studies of the same POWER sub-project, in a remote monitoring solution being developed by Altice Labs for Altice's SmartAL. In the appendices of this thesis, the API documentation of the final algorithms for the prediction module (Appendix A) and the recommendation system (Appendix B) are provided.

4.1 Prediction module

The methodology of the prediction module involves all the procedures related to data preparation and model training, testing and evaluation.

4.1.1 Dataset

All the algorithms are trained using data from the MyHeart dataset, currently available to the team. Although the MIMIC-III is also available, the MyHeart is considered more suitable for this study, since it contains home measured data rather than data from patients admitted to intensive care units, as is the case of MIMIC-III. The dataset contains data from 41 patients over 60 days, for multiple variables, including BP, weight, ECG, respiratory rate and bioimpedance. Each variable was registered once daily. The variable of interest in this study is BP, that in this dataset contains 2460 measurements total, 829 of which are missing values. Even though it is not specified what type of measurement the available BP variable represents, the range of values from minimum to maximum indicates that it may refer to SBP, since the lowest value is of 86.0 and the highest is of 197.0 mmHg. These peak values represent SBP categories from optimal to grade 3 hypertension, following the information on table 2.1. The mean value of 119.7 mmHg calculated for the variable supports this assumption, representing an optimal SBP value.

4.1.2 Pre-processing

Data pre-processing is a crucial step when dealing with datasets, since it involves data cleaning, transformation, and reduction for elimination of inconsistencies in the data. The development of a proper method for data pre-processing to be applied on this dataset is part of another study developed on the present POWER sub-project. However, since both studies were being developed simultaneously, basic data pre-processing was done to induce the necessary autonomy in this study until the implementation of the mentioned method. For that matter, and to maintain the complexity involving data treatment as reduced as possible, mean values were used as missing data replacements.

For each patient, the arithmetic mean ($\bar{x}$) of the available BP variable values (x_i) was computed ignoring present missing values:

$$\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N}, \tag{4.1}$$

where N represents the number of non-null values for each patient. Thereafter, missing values were located and replaced by the respective mean value.

4.1.3 Prediction models

In this phase, multiple models for BP prediction were developed using Python. Two libraries were used for the construction of the models' architectures: *Keras* with *TensorFlow* and *scikit-learn*. All the models considered the BP variable, available in the described dataset, as input.

Linear Regression

The first model to be considered was a simple least squares Linear Regression (LR) model that follows the premise stated in 2.6. This algorithm was the first choice because it allowed the assessment on whether BP evolution could be considered a linear problem. That way, if the LR model had a satisfactory performance, and there was no significant discrepancy between its results and the results of other models of higher complexity, the use of the latter would not be justifiable, and the prediction problem could be addressed through simpler linear algorithms.

Long Short-Term Memory

A LSTM model was also developed. The model's layer stack consisted of two LSTM layers, followed by one output layer. The first LSTM layer contains 50 units, while the remaining layers contain only 1 each. The stacking of two LSTM layers in this case, and the increased number of units in the first one, had the purpose of helping the model capture more complex patterns, which may be useful when dealing with time series with sophisticated dynamics, such as BP. Both LSTM layers make use of the *ReLu* activation function and the output layer has a linear activation function applied. The choice of the *ReLu* activation function was because it is one of the most computationally efficient functions, since it does not rely on complex calculations, and does not cause the vanishing gradient effect. Model compilation relied on the *Adam* optimizer function, generally recommended as the default optimizer for its fast computation time and few parameters, and mean squared error loss function, useful in preventing large errors for outlier prediction.

Since LSTMs are vastly used in state of the art studies, this network was developed as the simple nonlinear model for comparison with the other models.

Jump Neural Network

A JNN model similar to the one proposed in [39] for glucose prediction was considered. As described in 2.4.2, this network includes both a linear and nonlinear component in the model. For that, using layers from the mentioned libraries, the architecture was established through a dense hidden layer with a sigmoid activation function (nonlinear component), chosen for comparison with the *ReLu* in LSTM, that was then concatenated with the input layer. This allows for the input data to be connected directly to the dense output layer with a linear activation function (jump connection). The hidden layer was composed of 4 neurons and the output layer of only 1. The reduction in the number of neurons in the hidden layer, comparatively to LSTM, was to evaluate how well the complex dynamics of the signal were captured by the junction of both linear and nonlinear components and if the increased number of units in the previous model was justifiable. Model compilation relied on *Adam* optimizer function and mean squared error loss function, for the reasons previously described. A detailed description of this network can be found in the Appendix in [40].

This model is found useful in the present context because, as referred in [39], its structure deals with both the linear and nonlinear dynamics of time series signals, as is the case of BP signals. It will also serve as an alternative nonlinear model that can be compared to the usual LSTM approach.

Case-Based Reasoning

The developed CBR model is an adaptation of one that has been recently developed by the research team for predicting the evolution of glucose signals [47]. Given the reliable results achieved with its application to glucose prediction, the team proposed its further adaptation to BP prediction. As mentioned, this algorithm can retrieve known cases and propose a solution based on the similarity between known cases and the new case. Similarity assessment was performed using distance functions (e.g., *euclidean*, *mahalanobis*, *cityblock*, etc.), and a chosen number of the most similar cases were retrieved. The prediction results, here referred as solutions, were proposed through averaging or weighted averaging of the previously existing solutions of known similar cases.

4.1.4 Model development

Training process was done the same for LR, LSTM and JNN models. A leave-one-out cross-validation approach was considered by iterating the whole dataset and selecting one test patient at a time. In each iteration, the rest of the dataset was used for model training. Data from the various patients in the training dataset was separated into multiple N-length arrays, being N the number of days used as input, using a sliding window of size N and step 1. With a defined prediction horizon, P, two types of prediction were considered: single day prediction, P days after the input data, and multiple days prediction, in which all values up to P days after the input were predicted, that is, a P-length array. Regardless of the prediction type, each N-length input array was associated with a single value output. In the first case, the correspondent output was the P-th day after the input sequence. For the second case, the correspondent output was the value measured in the day that followed the input sequence. Input-output pairs were then used to fit the models. This data preparation process for a single day prediction, with $N = 4$ and $P = 3$ is pictured in figure 4.1.

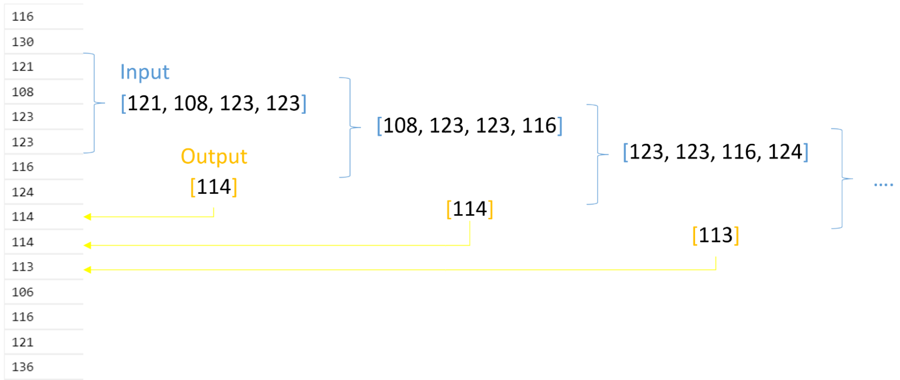

Figure 4.1: Example of the data preparation process for a single day prediction with $N = 4$ and $P = 3$.

A similar data preparation process was used for test patients' data. That way, the fitted model used each N-length array available as input for prediction. While for single day prediction the model only predicted the P-th day measurement value, for multiple day prediction, the measurement immediately after the input sequence was predicted and added to the input array to predict the next value, repeating this process until an array of all P days' values was obtained. For multiple day prediction, separate models could have been trained for each day of the prediction horizon to avoid using predicted values in the prediction of subsequent days. However, the use of the model that predicts the next day was carried out so that the complexity of the approach would not increase exponentially, with the creation of up to 9 models for each test patient. Even if this approach is prone to error accumulation, the creation of a model for next day prediction will theoretically have greater accuracy than models created for greater horizons and allow a comparison to be established with single day prediction. Each predicted value or array was then compared to the real test data values and multiple evaluation metrics, that will be addressed later in 4.1.6, were assessed. This process was repeated for all test patients and a mean value of each evaluation metric was calculated for each (N, P) pair model, by averaging of the results obtained from all test patients for the respective metrics.

As for the CBR model, a similar leave-one-out cross-validation approach was used. In each iteration, test patient's data was compared to the remaining patients' data, being the data from the training patients used as the case base knowledge. This data was divided into $(problem, solution)$ pairs, in which the problem was constituted by all the data up to a time and its solution by the remaining data, being the solution length equivalent to the prediction horizon P. The problems from the test patients were used for comparison with the training problems, and a solution was proposed for each. As in the previous cases, up to P blood pressure values were predicted for each test problem and, by comparison with the real solution of the respective test patient, the same evaluation metrics mentioned in 4.1.6 were computed. This process was once again considered for every test patient and the results represent the mean of each metric.

4.1.5 Hyperparameters tuning

To find the optimal hyperparameter values for the given algorithms, a grid search technique was applied. The hyperparameters in focus for LR, LSTM and JNN were both the input size N, and the prediction horizon P. In agreement with the project partners, given the telemonitoring context of this study, it was recommended that the input should try to rely on a small amount of data, since it will depend on manual data entry by the patient, which can lead to inconsistent measurement frequencies. Also, for the prediction horizon, the goal would be the prediction of a few days, around a week, after the input since this is a clinically reasonable horizon for predicting this type of physiological signals. Theoretically, the model will perform better the larger the amount of input data and the shorter the prediction horizon. However, under these conditions, there may also be cases of overfitting. That said, the models were trained for input sizes of $N = [4, 6, 8, 10]$ days and prediction horizon of $P = [3, 5, 7, 9]$ days, and subsequently evaluated.

In the case of the CBR model, the input size was constant for each prediction horizon, since all the training patients' data were used as the knowledge base problems, except for the last P values of each patient, that were the corresponding solutions in each case. That said, the hyperparameters that were submitted for tuning were the prediction horizon (in which the same values as before were considered, $P = [3, 5, 7, 9]$, in a similar approach), the distance metric used for similarity assessment, the number of similar cases retrieved to propose a solution for a new problem, and the type of case adaptation by which known solutions are adapted to propose a solution for a new problem.

4.1.6 Performance evaluation

To assess the models' performance, the predicted values ($\hat{x}$) were compared to the real measurements (x) through four evaluation metrics: Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The choice of these metrics was based on an assessment of the most widely used in the state of the art for regression problems of the sort.

MAPE is the measure of the average absolute percentage errors, showing the accuracy of the prediction in comparison to the actual measurement:

$$MAPE = \frac{1}{p}\sum_{j=1}^{p}\frac{|x_j - \hat{x}_j|}{|x_j|} \times 100. \tag{4.2}$$

MAE represents the average of the absolute error and measures the average of the residuals (difference between the observed and predicted values):

$$MAE = \frac{1}{p}\sum_{j=1}^{p}|x_j - \hat{x}_j|. \tag{4.3}$$

RMSE is the square root of the average of the squared errors between predicted and real values and computes the standard deviation of residuals:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{p}(x_j - \hat{x}_j)^2}{p}}. \tag{4.4}$$

In every equation, p represents the number of predicted and, consequently, real values to be compared. Besides MAPE that is expressed in percentage, the remaining metrics are expressed in the units of the compared values, i.e., mmHg in this case. The results of the evaluations are later presented in chapter 5.

4.1.7 Final model

Based on the evaluation of all the developed models, the model with overall best performance was selected and implemented as the prediction model. The selected model was JNN. The goal was to extend the selected model to a multivariate setting by incorporating discrete information that may be useful in BP prediction and implement it in parallel with the recommendation system. For example, as seen in section 3.3.1, data related to meals and exercise could be associated with BP variation to improve prediction accuracy. New data variables were supposed to be extracted from the data collection study performed with the partners at CHUC. This approach, as well as performance evaluation results and model selection, will be discussed in the next chapter.

As for the final implementation, previously trained JNN models were saved to be easily accessible for prediction. Even though some pairs of input size and prediction horizon displayed better performances, it was accorded that the model should support various data input sizes and prediction horizons to improve the flexibility of the application. It was also agreed with the project partners that the output of the model should be the prediction of BP for the first, third and seventh day after input. That said, multiple JNN models were trained for single day prediction, one per input size and prediction horizon pair, making a total of 24 models, with 8 models trained for each prediction horizon (1, 3 or 7 days), all supporting input sizes from 3 to 10 days. All the models were saved to be subsequently loaded into the prediction algorithm, so that it is prepared to receive the patient's input data and is able to adjust to its length. The input data and the remaining parameters all go through a validation process. The aforementioned methods of the final algorithms for data validation and BP prediction are described in the API included in the Appendix A of this document.

4.2 Recommendation module

A knowledge-based recommendation system was also developed. The goal of this system is to evaluate user contextual data and give advice on lifestyle habits changes that will help control or prevent hypertension, acting in parallel with the prediction model.

The knowledge basis of the model is founded on a set of rules established through conditions imposed on patients' contextual data. These conditions result from an adaptation of the various guidelines presented in 3.3.1. At this stage, multiple variables besides BP are needed for background information, such as sex, age, BMI, waist circumference, SBP and DBP values, HR, daily sodium intake, existing comorbidities, and also, for the past 7 days, the daily duration and intensity of aerobic exercise, the days in which resistance exercise was performed, and daily alcohol consumption. The set of rules established are based on the following guidelines:

- BMI should be maintained between 18.5 and 24.9 kg/m2.
- Waist circumference no higher than 94 cm for men and 80 cm for women.
- Limiting sodium intake to no more than 2g per day.
- Daily practice of no less than 30 min of moderate intensity or 15 min of high intensity aerobic exercise on at least 5 days a week.
- Resistance training on 2 or more days a week.
- Alcohol consumption should be limited to 2 units daily and 14 weekly for men or 1.5 daily and 8 weekly for women, with 1 unit representing around 8g of pure alcohol.

The patient's input variables will be analyzed according to the conditions described above and the system will be able to suggest changes that need to be made on them. For this purpose, an algorithm that implemented the rules in an if-then format was developed and its operation, as well as input data validation, are described in the API provided in the Appendix B.

5

Results and Discussion

In this chapter, the results of the performance evaluation of the models described in the previous chapter are presented and discussed. As mentioned, four models were developed for BP prediction: LR, LSTM, JNN and CBR. Of the four, the one considered most suitable for use in the present scenario was selected and implemented in parallel with the recommendation system.

Prediction module

In theory, considering the known properties of the models and the time series problem at hand, one could assume some aspects of the performances before analyzing the results. Daily BP levels, even when measured at the same time and at the same conditions every day, are not expected to follow a clear pattern, since BP is continuously changing and depends on many factors. That said, values may present abrupt changes as quickly as in a single time step, causing great volatility in evolution trends, which can be difficult to follow with a linear model. On the contrary, nonlinear models such as LSTM can capture patterns in the input sequential data and consider them in the prediction of future values, discarding the need to find a linear relation between the input and the targets, which is likely to be a more suitable approach in this case, compared to LR. Furthermore, in this sense, with the JNN characteristics covered in 2.4.2, there is also the possibility for this model to further enhance the benefits of LSTM, as it allows a direct and simultaneous evaluation of both linear and nonlinear dynamics of BP signals.

In general, it is also possible to deduce that single day prediction may produce lower errors than multiple days prediction, since in the latter case, the prediction of future values is dependent on the use of previously predicted values, which results in the accumulation of error as the prediction advances from the first day to the $P-th$ day. It is also likely that models trained with higher input sizes (N) and lower prediction horizons (P) will have better results than pairs with opposite configurations, not only because the increased amount of input data gives more information on signal variability, but also because the inconstancy of BP, caused by the multiple factors that have an influence on it, hinders long term prediction of the variable.

LR, LSTM, and JNN

The results of the performance of the LR, LSTM and JNN for both single and multiple days prediction are summarized in tables 5.1 and 5.2, respectively.

Table 5.1: Results for single day prediction for the LSTM, JNN and LR models. N = input size, P = prediction horizon.

		MAPE				RMSE & MAE			
	N \ P	3	5	7	9	3	5	7	9
LR	4	17.11	13.22	15.45	18.36	19.57	15.14	17.65	20.96
	6	12.21	16.85	14.61	21.43	14.00	19.24	16.69	24.47
	8	16.37	16.08	16.60	24.64	18.70	18.36	18.97	28.11
	10	14.32	19.24	18.82	31.48	16.36	21.98	21.48	35.91
LSTM	4	12.63	9.12	11.97	13.24	14.73	10.29	13.74	14.94
	6	6.64	29.53	7.56	11.32	7.75	33.33	8.99	12.64
	8	25.06	7.79	12.10	9.46	27.89	9.00	13.35	10.72
	10	5.80	7.07	7.40	6.83	6.64	8.28	8.64	7.84
JNN	4	3.87	3.94	4.00	4.14	4.67	4.76	4.82	4.98
	6	3.68	3.76	3.93	4.03	4.45	4.55	4.76	4.86
	8	3.66	3.74	3.90	4.02	4.42	4.55	4.73	4.88
	10	3.64	3.75	3.91	3.98	4.41	4.57	4.75	4.83

Table 5.2: Results for multiple days prediction for the LSTM, JNN and LR models. N = input size, P = prediction horizon.

		MAPE				**RMSE**				**MAE**			
	N \ P	3	5	7	9	3	5	7	9	3	5	7	9
LR	4	15.89	16.72	16.70	16.65	18.86	19.89	19.90	19.87	18.21	19.15	19.13	19.07
	6	16.52	16.94	16.90	16.85	19.54	20.09	20.09	20.06	18.90	19.37	19.32	19.27
	8	11.70	12.39	12.69	13.29	14.13	15.02	15.42	16.17	13.42	14.20	14.55	15.23
	10	15.00	16.12	16.26	16.63	17.86	19.25	19.44	19.89	17.15	18.41	18.57	19.00
LSTM	4	11.67	18.94	14.64	14.06	14.26	22.34	17.53	17.47	13.44	20.79	16.44	16.28
	6	8.94	11.28	29.19	15.41	10.79	14.21	40.78	19.34	9.98	12.83	29.67	17.37
	8	5.84	5.60	5.90	6.54	7.61	7.59	8.06	10.02	6.85	6.59	6.95	8.04
	10	5.80	164.85	193.7	6.61	7.69	160.85	275.75	9.04	6.96	159.24	196.69	7.80
JNN	4	3.93	4.20	4.19	4.38	5.43	5.97	6.09	6.35	4.72	5.02	5.01	5.20
	6	3.64	3.88	4.00	4.09	5.14	5.62	5.87	6.06	4.40	4.66	4.79	4.89
	8	3.59	3.68	3.77	3.78	5.08	5.43	5.67	5.78	4.35	4.45	4.56	4.57
	10	3.64	3.72	3.71	3.96	5.15	5.50	5.61	5.96	4.40	4.51	4.49	4.76

From the results shown in both tables we can confirm some of the assumptions. For an instance, in both cases LR is the model with the poorest performance, and JNN performed best overall. It is also possible to see in some cases that the performance improved as the input size increased and worsened as the prediction horizon increased.

For single day prediction, RMSE presented the same results as MAE, since in this approach the prediction results consist of only one value, and by analyzing the equations 4.3 and 4.4, it is easy to see that these are equivalent for single value comparison. From table 5.1, we see that LR obtained MAPEs between 12.21 and 31.48% and RMSEs or MAEs from 14.00 to 35.91 mmHg, LSTM had values of MAPE from 5.80 to 29.53% and RMSE or MAE from 6.64 to 33.33 mmHg, and finally JNN presents MAPEs between 3.64 and 4.14% and RMSEs or MAEs between 4.41 and 4.98 mmHg, being considered the best model for single day prediction.

As for multiple days prediction, LR once again presented the overall worst performance of the three models, with MAPE ranging from 11.70 to 16.94 %, RMSE ranging from 14.13 to 20.09 mmHg and MAE ranging from 13.42 to 19.37%. LSTM performance evaluation resulted in values of MAPE from 5.60 to 193.7 %, RMSE from 7.59 to 275.75 mmHg and MAE from 6.59 to 196.69 mmHg. At last, JNN exhibits once more the lowest errors for all evaluation metrics, with MAPE varying from 3.59 to 4.38%, RMSE from 5.08 to 6.35 mmHg and MAE from 4.35 to 5.20 mmHg.

Especially in the case of LSTM for multiple days prediction, some remarkably high error values are noticeable, being discrepant with the remaining results. This may be due to the LSTM's propensity for data overfitting. Since LSTM networks are overly complex and involve multiple parameters, the development of an optimized model requires a cautious hyperparameter tuning. For example, the number of LSTM neurons is an important parameter that highly influences the predictive capability of the algorithm. With few units, the model is unlikely to correctly comprehend the structure of the data, while with too many units it is prone to overfitting. Also, the use of *ReLu* activation functions in both LSTM layers in conjunction with the *Adam* optimizer function could have been prejudicial to the neural network performance. The *Adam* function may suffer from a weight decay problem, since at each epoch, the function slightly reduces the value of the weights. When considering a high number of epochs, this decay can be accentuated, leading to the incidence of negative weights that can introduce negative inputs into the *ReLu* function, causing the problem of the dying *ReLu*, in which the neurons will only output values of 0. Given the results, we can assume that the chosen parameters for the construction of the LSTM model is not ideal for some N, P pairs, even though it performed well for others, and the fragility of the algorithm would probably require an optimization of its parameters for each case.

CBR

The results obtained by the CBR model are summarized in table 5.3.

Table 5.3: Best results for each prediction horizon for the CBR model. P = prediction horizon.

	P	**No. retrieved cases**	**Distance**	**Adaptation**	**MAPE**	**RMSE**	**MAE**
CBR	3	10	*jaccard*	*average*	4.15	4.94	6.03
	5	10	*jaccard*	*average*	4.21	4.99	6.24
	7	9	*cosine*	*average*	4.10	4.89	6.13
	9	10	*cosine*	*weighted average*	4.41	5.29	6.80

Presented results are the best ones obtained for each prediction horizon, because the variation of the multiple hyperparameters resulted in many combinations (the performance of a total of 264 combinations was evaluated), so it made no sense to display them all. The complete set of results obtained present MAPEs ranging from 4.10 to 5.82%, RMSEs ranging from 4.89 to 6.91 mmHg and MAE ranging from 5.75 to 8.49 mmHg. Overall, *cosine* and *jaccard* distance functions were associated with the best performances, followed by *chebyshev*. No relevant difference in results was noticed for different numbers of retrieved cases or adaptation functions. One would expect the model to perform better with less retrieved cases and with the use of weighted averaging to the detriment of simple averaging. Since cases are retrieved based on similarity to the case under review and the most similar cases are retrieved first, an increase in the number of retrieved cases implies the retrieval of less similar cases. All the retrieved cases are used to propose a solution to the new case, and using simple averaging causes the less similar cases to have the same influence as the most similar cases, since all the cases will weigh the same in the process. This may be pointed as a drawback for the use of the proposed CBR model. Another drawback of this approach is that it is based primarily on the comparison of values, which implies that information about the dynamics of the time series is not considered and outlier predictions are hindered. In addition, it requires the establishment of a reliable library of known cases before it can be used, making it very dependent on the existence of data at an early stage of development.

Final remarks and model selection

Focusing on the results for RMSE, a useful metric for evaluation of accuracy of the models, it can be stated that the overall best performing models have errors around 5 mmHg or less. Based on what was presented in table 2.1 and state of the art literature [42, 43, 59], and taking into account the classification intervals of the categories of SBP and the extent of its values, these results can be considered satisfactory, allowing the discrimination of different BP states. In terms of MAPE, a similar opinion can be drawn, since the lower its value, the better the predictive ability of the model. Given that scores below 4% were obtained, and assuming that the values of SBP normally range between 100 and 200 mmHg (considering the scope of normo- and hypertension), these can represent an average difference of about 4 to 8 mmHg between real and predicted values, being considered satisfactory for the purpose of this application. On the same note, the MAEs of around 5 mmHg represent good results, indicating a reduced average error of the residuals.

In agreement with what has been said, the JNN was selected as the best performing model. Even though there may be room for improvement in all the proposed models, no further parameters were assessed in the models. The goal for this module was to build an algorithm capable of predicting the future dynamics of BP, so it would not be justifiable to implement other models for the same purpose when the developed JNN proved to be a competent method for all configurations, showing satisfactory results in both single and multiple days prediction for the various prediction horizons, regardless of the input size.

Visual representation examples of single day prediction using the selected model, are depicted in figures 5.1 to 5.4. One example is shown for each prediction horizon considered, using the JNN models with the best performing input size. The predicted signals always start at day $N+P$, since the first N days are used as input and the first prediction value corresponds to the value of the $P-th$ day after the input.

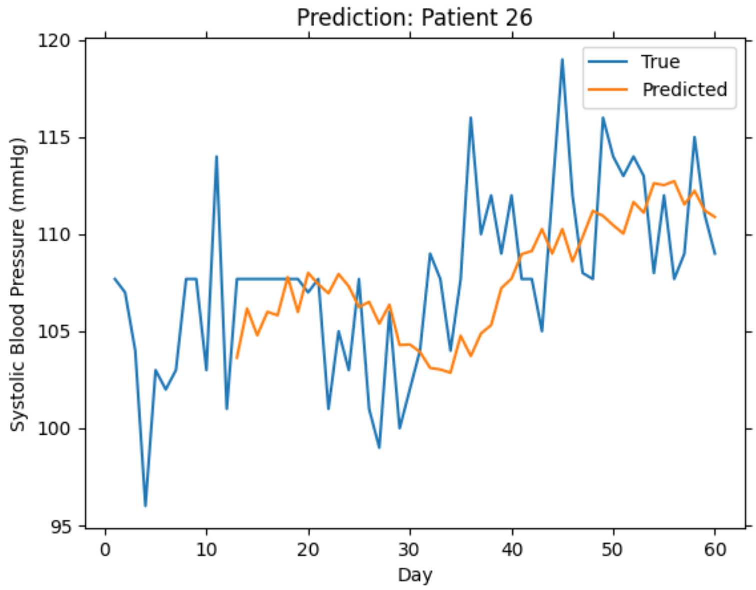

Figure 5.1: Visual representation of single day prediction results for patient 26 using JNN model with $N = 10$ and $P = 3$.

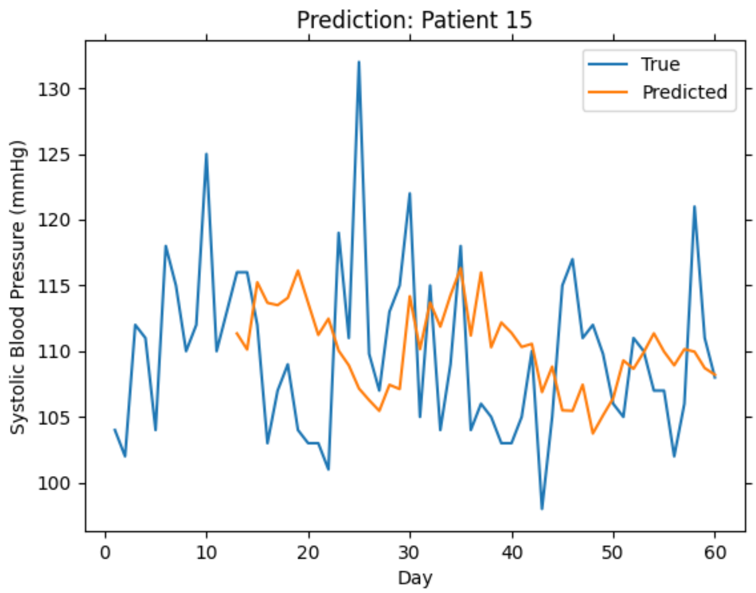

Figure 5.2: Visual representation of single day prediction results for patient 15 using JNN model with $N = 8$ and $P = 5$.

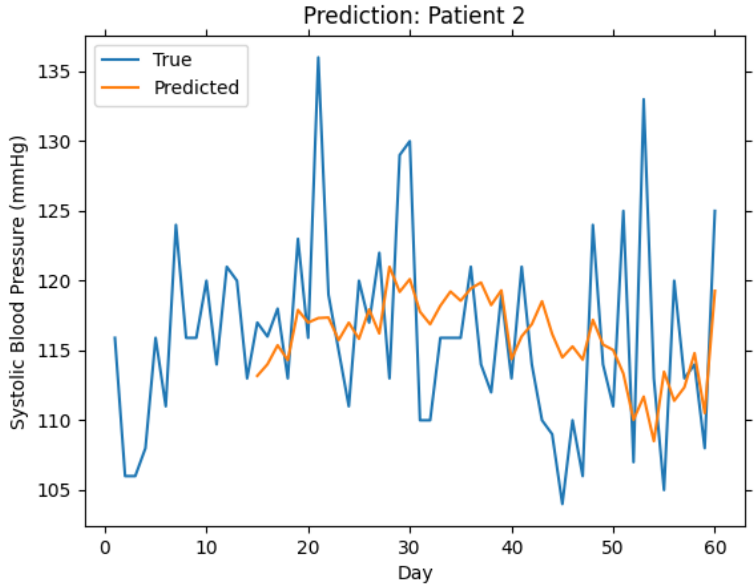

Figure 5.3: Visual representation of single day prediction results for patient 2 using JNN model with $N = 8$ and $P = 7$.

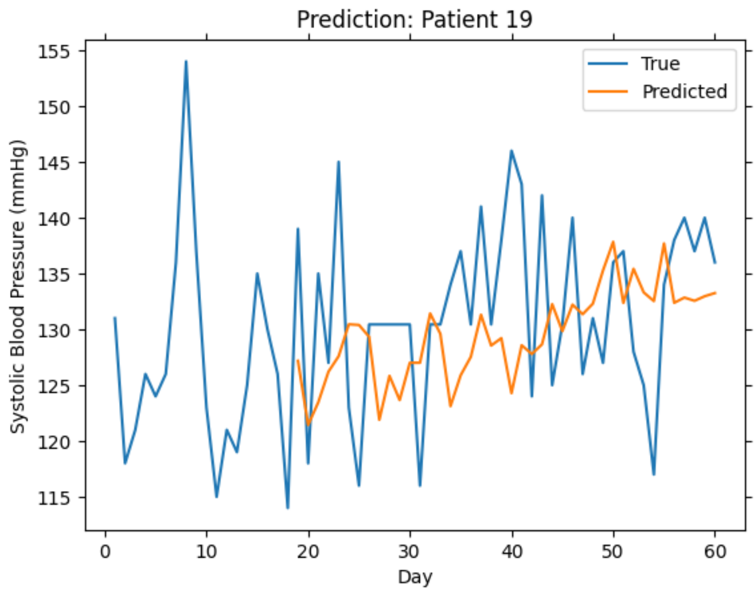

Figure 5.4: Visual representation of single day prediction results for patient 19 using JNN model with $N = 10$ and $P = 9$.

For multiple days prediction, graphs of the results are not presented in the same way, since each prediction step results in P values, which would lead to an overlap of values for the same days in the plot. However, example predictions of the P values, using the best performing multiple days JNN models for each prediction horizon are presented in figures 5.5 to 5.8.

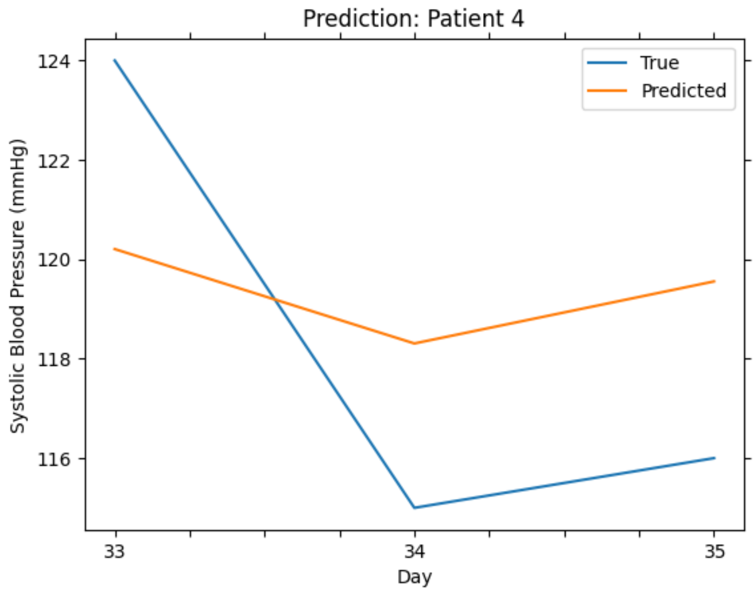

Figure 5.5: Visual representation of multiple days prediction results for patient 4 using JNN model with $N = 8$ and $P = 3$.

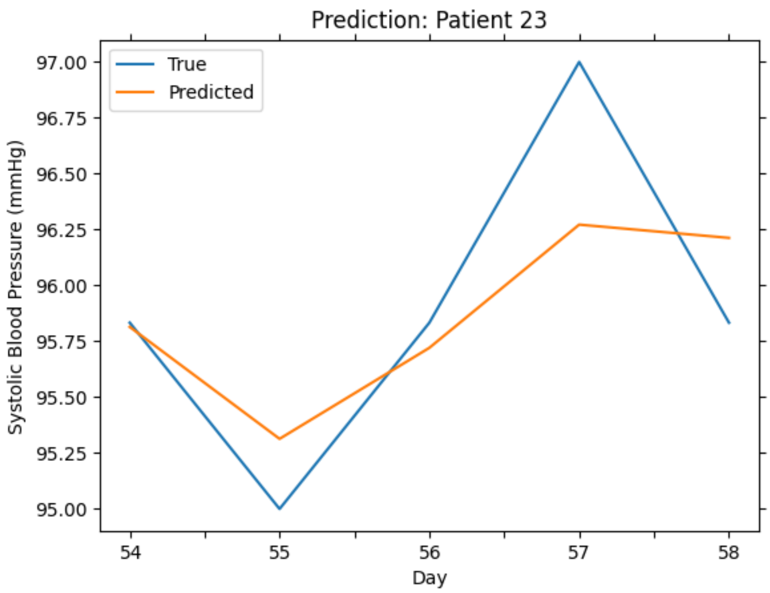

Figure 5.6: Visual representation of multiple days prediction results for patient 23 using JNN model with $N = 8$ and $P = 5$.

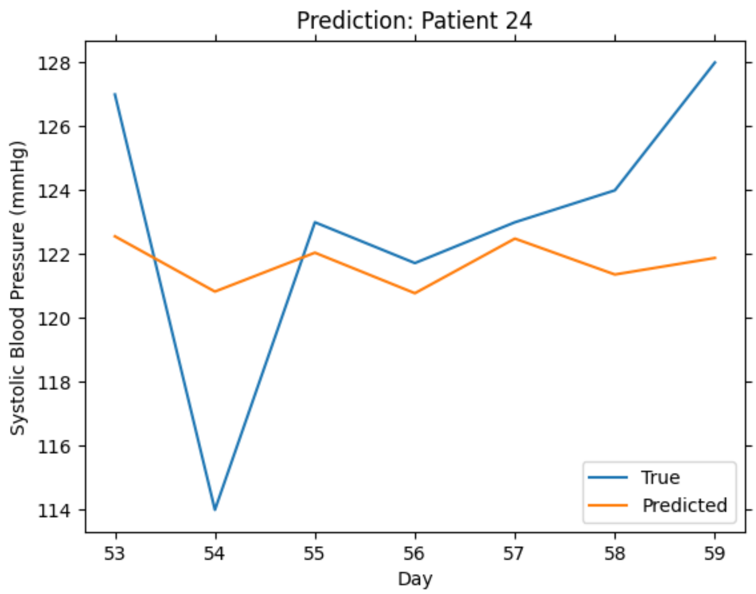

Figure 5.7: Visual representation of multiple days prediction results for patient 24 using JNN model with $N = 10$ and $P = 7$.

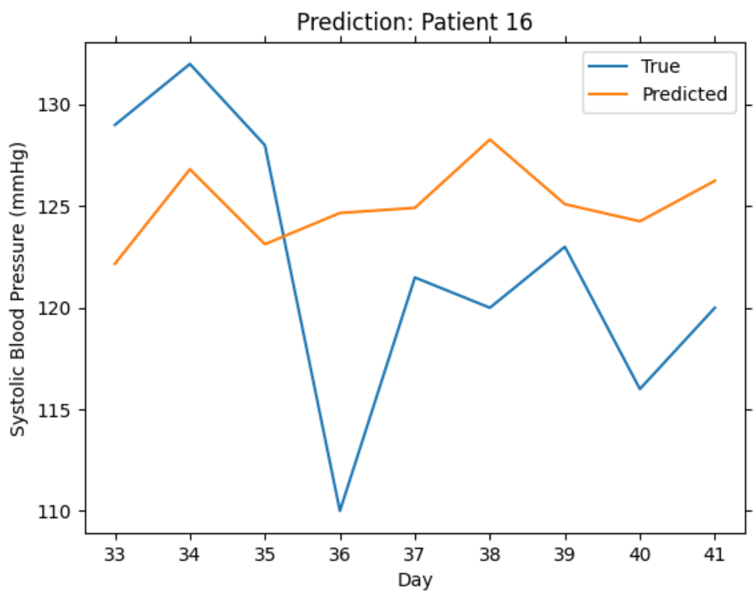

Figure 5.8: Visual representation of multiple days prediction results for patient 16 using JNN model with $N = 8$ and $P = 9$.

As seen in the figures, even if the predictions are not completely accurate, it is possible to verify that the predicted signal is able to approximately follow the evolution trends of the original signal.

Despite also achieving a satisfactory performance in multi-day prediction, with errors only slightly larger than the alternative, the single day prediction was chosen to meet the agreed output values for the first, third and seventh day after input. This could also have be achieved by a continuous multiple days prediction for the span of a week but since these are three individual single day approaches, it is not justified to follow that course of actions, knowing that it implies the prediction of values that are not necessary and may introduce accumulation of error in the process, decreasing the predictive accuracy.

The presented results can be considered reliable outcomes for the purpose study, given that the models only consider BP values. When it comes to the real application of the model, due to the complexity of the problem, further analysis is required. It is believed that the inclusion of other variables (e.g., related to meals and exercise) that are known to influence this physiological signal could benefit BP prediction and improve the overall accuracy of the model. For this, approaches similar to those described in some of the state of the art studies for treatment and inclusion of contextual data in the models could be considered.

Recommendation module

As for the recommendation model, the rules formed from the guidelines summarized in section 4.2 are schematized in table 5.4.

Table 5.4: Set of rules established using the available clinical guidelines. M = Male; F = Female

Rule ID	Input variables	Condition	Recommendation
R01	BMI	BMI < 18.5 kg/m2	INCREASE BMI
R02	BMI	BMI $\geq$ 25 kg/m2	DECREASE BMI
R03	Sex; Waist circumference	Sex = M and waist circumference $\geq$ 94 cm	DECREASE waist circumference
R04	Sex; Waist circumference	Sex = F and waist circumference $\geq$ 80 cm	DECREASE waist circumference
R05	Daily duration of aerobic exercise over the past week	Total weekly aerobic exercise duration $\leq$ 150 min	INCREASE daily aerobic exercise duration
R06	Daily duration of aerobic exercise over the past week	Daily aerobic duration = 0 min on 2 or more days	INCREASE daily aerobic exercise frequency
R07	Frequency of resistance exercise over the past week	Weekly frequency of resistance exercise < 2 days	INCREASE weekly resistance exercise frequency
R08	Daily sodium intake	Daily sodium intake > 2g	DESCREASE daily sodium intake
R09	Sex; Alcohol consumption over the past week	Sex = M and weekly alcohol consumption > 14 units	DECREASE weekly alcohol consumption
R10	Sex; Alcohol consumption over the past week	Sex = F and weekly alcohol consumption > 8 units	DECREASE weekly alcohol consumption
R11	Sex; Alcohol consumption over the past week	Sex = M and daily alcohol consumption > 2 units	DECREASE daily alcohol consumption
R12	Sex; Alcohol consumption over the past week	Sex = F and daily alcohol consumption > 1.5 units	DECREASE daily alcohol consumption

Rules from *R01* to *R04* recommend changes in BMI and waist circumference. Although the considered guidelines do not mention effective methods to directly reduce/increase these factors, they are both known to be related to body weight. Therefore, changes that influence body weight, such as physical exercise and diet, are expected to alter BMI and waist circumference. This way, the patient can be guided to follow the mentioned dietary and exercise recommendations for that purpose.

For better understanding of rules *R05* to *R07*, some remarks need to be made. In rule *R05*, given the indication from the guidelines that patients should incur in 30 min of moderate intensity or 15 min of high intensity aerobic exercise daily, having information about daily exercise duration and intensity, the model assumes that the duration of high intensity training will weigh twice as much as moderate intensity training. Thus, a total of at least 150 min per week should be performed.

In rule *R06*, the same inputs as in the previous rule will be analyzed to see if the patient practiced the recommended duration of aerobic exercise on at least 5 days a week. Rule *R07* only considers the recommendation that resistance exercise should be practiced at least 2 days a week, with no information about its duration or intensity.

There are other variables that are known to be related to variations in BP but have not been included in the rules due to lack of information regarding their direct influence on BP in the guidelines, that would allow conditions based on them to be established. For example, the user's existing comorbidities are a crucial factor to take into consideration, since recommendations that are beneficial in controlling hypertension may be detrimental for patients suffering simultaneously from other complications. That said, this set of rules was just a prototype built using only the available guidelines, with the intention of being presented to project partners at CHUC to be improved. Knowledge-based recommendation systems to be applied in clinical domain require high quality knowledge bases and, in this context, the support of health professionals with solid knowledge about chronic disease management and patient monitoring is helpful. Nevertheless, an initial version of a recommendation system was developed, as mentioned before. An example of the generation of a personalized recommendation according to one of the rules is shown below:

Example: **if** (sex == 'F' **and** weeklyAlcoholConsumption > 14) **or** (sex == 'M' **and** weeklyAlcoholConsumption > 8) **then** Recommendation (Decrease weekly alcohol consumption: total weekly alcohol consumption should be limited to 14 and 8 units for men and women, respectively, with one unit corresponding to around 8g alcohol.)

For the reasons described above, there are no further results to present for this module. A possible approach for the evaluation of such system would be to compare recommendations generated by the algorithm with recommendations suggested by a qualified health professional, to verify the relevance of the produced recommendations for a given set of input data.

6

Conclusion and Future work

The present study introduces the development of a blood pressure prediction model and a lifestyle recommendation system to be included in a remote monitoring solution for the management of patients with hypertension.

For the prediction of blood pressure, four models, both linear and nonlinear, were developed and evaluated for different settings of input data and prediction horizons. Despite the dynamic characteristics of the time variable in question, the achievement of models with satisfactory performances proved the feasibility of this approach and the potential of its application in clinical settings. A JNN model was selected as the best performing algorithm, with the results suggesting a reasonable prediction of blood pressure evolution trends and an overall outperformance over models presented in the literature.

Regarding the recommendation system, an initial set of rules has been developed based on clinical practice guidelines from globally recognized institutions and societies specialized in hypertension. An initial version of a knowledge-based recommendation system was implemented using an if-then conditions set. The system ultimately outputs a series of recommendations that encourage changes in lifestyle habits according to the comparison of the patient's contextual data with the previously established rules.

The main limitation in this study was the lack of easily accessible and reliable databases, with a vast amount of not only temporal data related to the necessary physiological factors but also of discrete contextual data that would allow the formation of a background on the patient's condition, for a better understanding of the evolution tendencies of the time series signal under evaluation.

For the prediction algorithm, the provision of contextual data will enable the use of collected data regarding factors that influence BP variation to form new features to be considered in the process (e.g., weight and discrete information related to exercise and meals), and permit the extension of the selected model to a multivariable setting. This could be a valuable addition to the prediction model, allowing a better capture of blood pressure dynamics and consequently a more accurate prediction and improved model performance.

In the case of the recommendation system, collaboration with healthcare professionals at CHUC will provide qualified insight into the influence of various physiological and lifestyle factors on blood pressure variations, thus permitting the improvement of the proposed rules and obtaining new rules to further extend the knowledge-based recommendation system.

In the future, a clinical data collection study will be initiated in collaboration with the partners at CHUC and Altice Labs, to complement data currently available to the team on private databases, which will allow the validation and improvement of the already developed modules and its posterior implementation in Altice's SmartAL telemonitoring platform for hypertension management.

Bibliography

[1] R. E. Klabunde, *Cardiovascular Physiology Concepts.* LWW, 2 ed., 2011.

[2] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, "Deep learning for time series forecasting: a survey," *Big Data*, vol. 9, no. 1, pp. 3–21, 2021.

[3] P. McNelis, *Neural Networks in Finance: Gaining Predictive Edge in the Market.* Academic Press Advanced Finance, Elsevier Science, 2005.

[4] C. Harpham and C. W. Dawson, "The effect of different basis functions on a radial basis function network for time series prediction: A comparative study," *Neurocomputing*, vol. 69, no. 16-18, pp. 2161–2170, 2006.

[5] A. Eliasy and J. Przychodzen, "The role of ai in capital structure to enhance corporate funding strategies," *Array*, vol. 6, p. 100017, 2020.

[6] B. Williams, G. Mancia, W. Spiering, E. Agabiti Rosei, M. Azizi, M. Burnier, D. L. Clement, A. Coca, G. de Simone, A. Dominiczak, T. Kahan, F. Mahfoud, J. Redon, L. Ruilope, A. Zanchetti, M. Kerins, S. E. Kjeldsen, R. Kreutz, S. Laurent, G. Y. H. Lip, R. McManus, K. Narkiewicz, F. Ruschitzka, R. E. Schmieder, E. Shlyakhto, C. Tsioufis, V. Aboyans, I. Desormais, and ESC Scientific Document Group, "2018 ESC/ESH guidelines for the management of arterial hypertension," *Eur. Heart J.*, vol. 39, pp. 3021–3104, Sept. 2018.

[7] WHO, "Hypertension," 2021. Accessed: 2022-06-21.

[8] J. G. F. Cleland, A. A. Louis, A. S. Rigby, U. Janssens, A. H. M. M. Balk, and TEN-HMS Investigators, "Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care management system (TEN-HMS) study," *J. Am. Coll. Cardiol.*, vol. 45, pp. 1654–1664, May 2005.

[9] K. L. Tucker, J. P. Sheppard, R. Stevens, H. B. Bosworth, A. Bove, E. P. Bray, K. Earle, J. George, M. Godwin, B. B. Green, *et al.*, "Self-monitoring

of blood pressure in hypertension: a systematic review and individual patient data meta-analysis," *PLoS medicine*, vol. 14, no. 9, p. e1002389, 2017.

[10] S. Omboni and A. Guarda, "Impact of home blood pressure telemonitoring and blood pressure control: a meta-analysis of randomized controlled studies," *American journal of hypertension*, vol. 24, no. 9, pp. 989–998, 2011.

[11] A. M. Ward, O. Takahashi, R. Stevens, and C. Heneghan, "Home measurement of blood pressure and cardiovascular disease: systematic review and meta-analysis of prospective studies," *Journal of hypertension*, vol. 30, no. 3, pp. 449–456, 2012.

[12] K. L. Tucker, J. P. Sheppard, R. Stevens, H. B. Bosworth, A. Bove, E. P. Bray, M. Godwin, B. Green, P. Hebert, F. D. R. Hobbs, I. Kantola, S. Kerry, D. J. Magid, J. Mant, K. L. Margolis, B. McKinstry, S. Omboni, O. Ogedegbe, G. Parati, N. Qamar, J. Varis, W. Verberk, B. J. Wakefield, and R. J. McManus, "Individual patient data meta-analysis of self-monitoring of blood pressure (BP-SMART): a protocol," *BMJ Open*, vol. 5, p. e008532, Sept. 2015.

[13] J. J. Boyne and H. J. Vrijhoef, "Implementing telemonitoring in heart failure care: barriers from the perspectives of patients, healthcare professionals and healthcare organizations," *Current heart failure reports*, vol. 10, no. 3, pp. 254–261, 2013.

[14] A. C. Guyton, J. E. Hall, *et al.*, *Textbook of medical physiology*, vol. 548. Saunders Philadelphia, 1986.

[15] N. Herring and D. J. Paterson, *Levick's introduction to cardiovascular physiology.* CRC Press, 2018.

[16] J. Pumprla, K. Howorka, D. Groves, M. Chester, and J. Nolan, "Functional assessment of heart rate variability: physiological basis and practical applications," *International journal of cardiology*, vol. 84, no. 1, pp. 1–14, 2002.

[17] A. Roy, W. C. Fields, C. Rocha-Resende, R. R. Resende, S. Guatimosim, V. F. Prado, R. Gros, and M. A. Prado, "Cardiomyocyte-secreted acetylcholine is required for maintenance of homeostasis in the heart," *The FASEB Journal*, vol. 27, no. 12, p. 5072, 2013.

[18] S. Reule and P. E. Drawz, "Heart rate and blood pressure: any possible implications for management of hypertension?," *Current hypertension reports*, vol. 14, no. 6, pp. 478–484, 2012.

[19] AHA, "Blood pressure vs. heart rate (pulse)," Apr 2022. Accessed: 2022-08-29.

[20] J. Lapum, M. Verkuyl, W. Garcia, O. St-Amant, and A. Tan, *Vital Sign Measurement Across the Lifespan - 1st Canadian Edition*. 2019.

[21] G. Beevers, G. Y. Lip, and E. O'Brien, "ABC of hypertension: The pathophysiology of hypertension," *BMJ*, vol. 322, pp. 912–916, Apr. 2001.

[22] U. o. M. Healthwise Staff, "Secondary high blood pressure," 2020. Accessed: 2022-06-21.

[23] S. Oparil, M. C. Acelajado, G. L. Bakris, D. R. Berlowitz, R. Cífková, A. F. Dominiczak, G. Grassi, J. Jordan, N. R. Poulter, A. Rodgers, and P. K. Whelton, "Hypertension," *Nat. Rev. Dis. Primers*, vol. 4, p. 18014, Mar. 2018.

[24] R. H. Fagard and V. A. Cornelissen, "Effect of exercise on blood pressure control in hypertensive patients," *Eur. J. Cardiovasc. Prev. Rehabil.*, vol. 14, pp. 12–17, Feb. 2007.

[25] K. M. Diaz and D. Shimbo, "Physical activity and the prevention of hypertension," *Current hypertension reports*, vol. 15, no. 6, pp. 659–668, 2013.

[26] M. F. Leitzmann, Y. Park, A. Blair, R. Ballard-Barbash, T. Mouw, A. R. Hollenbeck, and A. Schatzkin, "Physical activity recommendations and decreased risk of mortality," *Archives of internal medicine*, vol. 167, no. 22, pp. 2453–2460, 2007.

[27] V. A. Cornelissen and N. A. Smart, "Exercise training for blood pressure: a systematic review and meta-analysis," *Journal of the American heart association*, vol. 2, no. 1, p. e004473, 2013.

[28] M. d. F. Monteiro and D. C. Sobral Filho, "Physical exercise and blood pressure control," *Revista Brasileira de Medicina do Esporte*, vol. 10, pp. 513–516, 2004.

[29] J. Jordan, V. Yumuk, M. Schlaich, P. M. Nilsson, B. Zahorska-Markiewicz, G. Grassi, R. E. Schmieder, S. Engeli, and N. Finer, "Joint statement of the european association for the study of obesity and the european society of hypertension: obesity and difficult to treat arterial hypertension," *J. Hypertens.*, vol. 30, pp. 1047–1055, June 2012.

[30] D. W. Harsha and G. A. Bray, "Weight loss and blood pressure control (pro)," *Hypertension*, vol. 51, pp. 1420–5; discussion 1425, June 2008.

[31] S. Tasnim, C. Tang, V. M. Musini, and J. M. Wright, "Effect of alcohol on blood pressure," *Cochrane Libr.*, vol. 2020, July 2020.

[32] P. Primatesta, E. Falaschetti, S. Gupta, M. G. Marmot, and N. R. Poulter, "Association between smoking and blood pressure: evidence from the health survey for england," *Hypertension*, vol. 37, pp. 187–193, Feb. 2001.

[33] M. V. Rainforth, R. H. Schneider, S. I. Nidich, C. Gaylord-King, J. W. Salerno, and J. W. Anderson, "Stress reduction programs in patients with elevated blood pressure: a systematic review and meta-analysis," *Current hypertension reports*, vol. 9, no. 6, pp. 520–528, 2007.

[34] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting.* Springer, 2002.

[35] C. Chatfield, *Time-series forecasting.* Chapman and Hall/CRC, 2000.

[36] R. J. Frank, N. Davey, and S. P. Hunt, "Time series prediction and neural networks," *Journal of intelligent and robotic systems*, vol. 31, no. 1, pp. 91–103, 2001.

[37] M. Kuhn, K. Johnson, *et al.*, *Applied predictive modeling*, vol. 26. Springer, 2013.

[38] P. S. Cowpertwait and A. V. Metcalfe, *Introductory time series with R.* Springer Science & Business Media, 2009.

[39] C. Zecchin, A. Facchinetti, G. Sparacino, and C. Cobelli, "Jump neural network for online short-time prediction of blood glucose from continuous monitoring sensors and meal information," *Computer methods and programs in biomedicine*, vol. 113, no. 1, pp. 144–152, 2014.

[40] C. Zecchin, A. Facchinetti, G. Sparacino, and C. Cobelli, "How much is short-term glucose prediction in type 1 diabetes improved by adding insulin delivery and meal content information to CGM data? a proof-of-concept study," *J. Diabetes Sci. Technol.*, vol. 10, pp. 1149–1160, Sept. 2016.

[41] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural computation*, vol. 3, no. 2, pp. 246–257, 1991.

[42] X. Li, S. Wu, and L. Wang, "Blood pressure prediction via recurrent models with contextual layer," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 685–693, 2017.

[43] J. Liu, Y. Wu, Z. Yuan, and X. Sun, “Blood pressure prediction with multi-cue based RBF and LSTM model,” in *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, IEEE, Oct. 2018.

[44] S. Masum, J. P. Chiverton, Y. Liu, and B. Vuksanovic, “Investigation of machine learning techniques in forecasting of blood pressure time series data,” in *Lecture Notes in Computer Science*, Lecture notes in computer science, pp. 269–282, Cham: Springer International Publishing, 2019.

[45] P. T. Yamak, L. Yujian, and P. K. Gadosey, “A comparison between arima, lstm, and gru for time series forecasting,” in *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, pp. 49–55, 2019.

[46] J. Kolodner, *Case-based reasoning.* Morgan Kaufmann, 2014.

[47] S. Zulj, P. Carvalho, R. T. Ribeiro, R. Andrade, and R. Magjarevic, “Data size considerations and hyperparameter choices in case-based reasoning approach to glucose prediction,” *Biocybernetics and Biomedical Engineering*, vol. 41, no. 2, pp. 733–745, 2021.

[48] G. Paré, K. Moqadem, G. Pineau, and C. St-Hilaire, “Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review,” *J. Med. Internet Res.*, vol. 12, p. e21, June 2010.

[49] G. Paré, M. Jaana, and C. Sicotte, “Systematic review of home telemonitoring for chronic diseases: the evidence base,” *J. Am. Med. Inform. Assoc.*, vol. 14, pp. 269–277, May 2007.

[50] S. Omboni, “Telemedicine for hypertension management: where we stand, where we are headed,” *Conn Health*, vol. 1, no. 2, pp. 85–97, 2022.

[51] R. Agarwal, J. E. Bills, T. J. W. Hecht, and R. P. Light, “Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control: a systematic review and meta-analysis,” *Hypertension*, vol. 57, pp. 29–38, Jan. 2011.

[52] P. K. Whelton, R. M. Carey, W. S. Aronow, D. E. Casey, Jr, K. J. Collins, C. Dennison Himmelfarb, S. M. DePalma, S. Gidding, K. A. Jamerson, D. W. Jones, E. J. MacLaughlin, P. Muntner, B. Ovbiagele, S. C. Smith, Jr, C. C. Spencer, R. S. Stafford, S. J. Taler, R. J. Thomas, K. A. Williams, Sr, J. D. Williamson, and J. T. Wright, Jr, “2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA

guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the american college of cardiology/american heart association task force on clinical practice guidelines," *Circulation*, vol. 138, pp. e484–e594, Oct. 2018.

[53] S. Omboni, R. J. McManus, H. B. Bosworth, L. C. Chappell, B. B. Green, K. Kario, A. G. Logan, D. J. Magid, B. Mckinstry, K. L. Margolis, G. Parati, and B. J. Wakefield, "Evidence and recommendations on the use of telemedicine for the management of arterial hypertension: An international expert position paper," *Hypertension*, vol. 76, pp. 1368–1383, Nov. 2020.

[54] Y. Duan, Z. Xie, F. Dong, Z. Wu, Z. Lin, N. Sun, and J. Xu, "Effectiveness of home blood pressure telemonitoring: a systematic review and meta-analysis of randomised controlled studies," *J. Hum. Hypertens.*, vol. 31, pp. 427–437, July 2017.

[55] W. J. Verberk, A. G. H. Kessels, and T. Thien, "Telecare is a valuable tool for hypertension management, a systematic review and meta-analysis," *Blood Press. Monit.*, vol. 16, pp. 149–155, June 2011.

[56] S. Omboni and R. Ferrari, "The role of telemedicine in hypertension management: focus on blood pressure telemonitoring," *Curr. Hypertens. Rep.*, vol. 17, p. 535, Apr. 2015.

[57] L. E. Burke, J. Ma, K. M. J. Azar, G. G. Bennett, E. D. Peterson, Y. Zheng, W. Riley, J. Stephens, S. H. Shah, B. Suffoletto, T. N. Turan, B. Spring, J. Steinberger, C. C. Quinn, and American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and Stroke Council, "Current science on consumer use of mobile health for cardiovascular disease prevention: A scientific statement from the american heart association," *Circulation*, vol. 132, pp. 1157–1213, Sept. 2015.

[58] G. Parati and S. Omboni, "Role of home blood pressure telemonitoring in hypertension management: an update," *Blood Press. Monit.*, vol. 15, pp. 285–295, Dec. 2010.

[59] H. Koshimizu, R. Kojima, K. Kario, and Y. Okuno, "Prediction of blood pres-

sure variability using deep neural networks," *Int. J. Med. Inform.*, vol. 136, p. 104067, Apr. 2020.

[60] T. Rocha, S. Paredes, P. De Carvalho, and J. Henriques, "Prediction of acute hypotensive episodes by means of neural network multi-models," *Computers in biology and medicine*, vol. 41, no. 10, pp. 881–890, 2011.

[61] T. Fujiwara, S. Hoshide, H. Kanegae, M. Nishizawa, and K. Kario, "Reliability of morning, before-dinner, and at-bedtime home blood pressure measurements in patients with hypertension," *J. Clin. Hypertens. (Greenwich)*, vol. 20, pp. 315–323, Feb. 2018.

[62] NHS, "About furosemide," Feb 2022. Accessed: 2022-08-04.

[63] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, p. 160035, May 2016.

[64] MIT-LCP, "Mimic documentation." Accessed: 2022-08-05.

[65] J. Habetha, "The myheart project-fighting cardiovascular diseases by prevention and early diagnosis," in *2006 international conference of the IEEE engineering in medicine and biology society*, pp. 6746–6749, IEEE, 2006.

[66] A. Gacek and W. Pedrycz, *ECG signal processing, classification and interpretation: a comprehensive framework of computational intelligence*. Springer Science & Business Media, 2011.

[67] T. J. Niiranen, L. Thijs, K. Asayama, J. K. Johansson, T. Ohkubo, M. Kikuya, J. Boggia, A. Hozawa, E. Sandoya, G. S. Stergiou, I. Tsuji, A. M. Jula, Y. Imai, J. A. Staessen, and IDHOCO Investigators, "The international database of HOme blood pressure in relation to cardiovascular outcome (IDHOCO): moving from baseline characteristics to research perspectives," *Hypertens. Res.*, vol. 35, pp. 1072–1079, Nov. 2012.

[68] T. J. Niiranen, K. Asayama, L. Thijs, J. K. Johansson, T. Ohkubo, M. Kikuya, J. Boggia, A. Hozawa, E. Sandoya, G. S. Stergiou, I. Tsuji, A. M. Jula, Y. Imai, J. A. Staessen, and International Database of Home blood pressure in relation to Cardiovascular Outcome Investigators, "Outcome-driven thresholds for home blood pressure measurement: international database of home blood pressure in relation to cardiovascular outcome," *Hypertension*, vol. 61, pp. 27–34, Jan. 2013.

[69] G. S. Stergiou, K. Asayama, L. Thijs, A. Kollias, T. J. Niiranen, A. Hozawa, J. Boggia, J. K. Johansson, T. Ohkubo, I. Tsuji, A. M. Jula, Y. Imai, J. A. Staessen, and International Database on HOme blood pressure in relation to Cardiovascular Outcome (IDHOCO) Investigators, "Prognosis of white-coat and masked hypertension: International database of HOme blood pressure in relation to cardiovascular outcome," *Hypertension*, vol. 63, pp. 675–682, Apr. 2014.

[70] J. A. Kotecha, D. Manca, A. Lambert-Lanning, K. Keshavjee, N. Drummond, M. Godwin, M. Greiver, W. Putnam, M.-T. Lussier, and R. Birtwhistle, "Ethics and privacy issues of a practice-based surveillance system: Need for a national-level institutional research ethics board and consent standards," *Canadian Family Physician*, vol. 57, no. 10, pp. 1165–1173, 2011.

[71] S. Garies, R. Birtwhistle, N. Drummond, J. Queenan, and T. Williamson, "Data resource profile: national electronic medical record data from the canadian primary care sentinel surveillance network (cpcssn)," *International Journal of Epidemiology*, vol. 46, no. 4, pp. 1091–1092f, 2017.

[72] T. Unger, C. Borghi, F. Charchar, N. A. Khan, N. R. Poulter, D. Prabhakaran, A. Ramirez, M. Schlaich, G. S. Stergiou, M. Tomaszewski, R. D. Wainford, B. Williams, and A. E. Schutte, "2020 international society of hypertension global hypertension practice guidelines," *Hypertension*, vol. 75, pp. 1334–1357, June 2020.

[73] R. S. Blumenthal and M. J. Blaha, "High blood pressure: Prevention, treatment and research," Aug 2021. Accessed:2022-08-18.

[74] S. Umemura, H. Arima, S. Arima, K. Asayama, Y. Dohi, Y. Hirooka, T. Horio, S. Hoshide, S. Ikeda, T. Ishimitsu, M. Ito, S. Ito, Y. Iwashima, H. Kai, K. Kamide, Y. Kanno, N. Kashihara, Y. Kawano, T. Kikuchi, K. Kitamura, T. Kitazono, K. Kohara, M. Kudo, H. Kumagai, K. Matsumura, H. Matsuura, K. Miura, M. Mukoyama, S. Nakamura, T. Ohkubo, Y. Ohya, T. Okura, H. Rakugi, S. Saitoh, H. Shibata, T. Shimosawa, H. Suzuki, S. Takahashi, K. Tamura, H. Tomiyama, T. Tsuchihashi, S. Ueda, Y. Uehara, H. Urata, and N. Hirawa, "The japanese society of hypertension guidelines for the management of hypertension (JSH 2019)," *Hypertens. Res.*, vol. 42, pp. 1235–1481, Sept. 2019.

[75] V. Vives-Boix, D. Ruiz-Fernández, A. d. Ramón-Fernández, D. Marcos-Jorquera, and V. Gilart-Iglesias, "A knowledge-based clinical decision support

system for monitoring chronic patients," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pp. 435–443, Springer, 2017.

[76] Z. Wang, H. Huang, L. Cui, J. Chen, J. An, H. Duan, H. Ge, N. Deng, *et al.*, "Using natural language processing techniques to provide personalized educational materials for chronic disease patients in china: development and assessment of a knowledge-based health recommender system," *JMIR medical informatics*, vol. 8, no. 4, p. e17642, 2020.

[77] D. V. Silveira, M. S. Marcolino, E. L. Machado, C. G. Ferreira, M. B. M. Alkmim, E. S. Resende, B. C. Carvalho, A. P. Antunes, and A. L. P. Ribeiro, "Development and evaluation of a mobile decision support system for hypertension management in the primary care setting in brazil: mixed-methods field study on usability, feasibility, and utility," *JMIR mHealth and uHealth*, vol. 7, no. 3, p. e9869, 2019.

[78] E. Charles, D. Stanley, and E. Agbaeze, "Knowledge-based diet and physical exercise advisory system," *International Journal of Science and Research (IJSR)*, vol. 14, no. 7, pp. 2319–7064, 2013.

Appendices

A

Prediction module API

POWER

empowering a digital future

DEFINIÇÃO DA INTERFACE DO MÓDULO DE PREDIÇÃO DE PRESSÃO ARTERIAL

Listagem de métodos

- predict_bp
- validate_bp_prediction_data

predict_bp

predict_bp(patient_input, pred_horizon)

Retorna um um *array* de tamanho 3. Quando os parâmetros são inválidos o *array* atribui todos os valores -1. Quando os parâmetros se encontram válidos o *array* é representado na primeira posição pelo valor da previsão correspondente a 1 dia após o último valor de input, na segunda posição o valor da previsão correspondente a 3 dias após o último valor de input e na última posição o valor da previsão correspondente a 7 dias após o último valor de input.

Parameters

patient_input : array_like

Array com dados de input. 1 valor de pressão arterial sistólica por dia. 3 a 10 valores, que devem ser inteiros ou *float*.

pred_horizon: None or int

O argumento do horizonte de previsão pode estar vazio (*None*) ou pode ser um número inteiro: 1, 3 ou 7 (correspondente a 1 dia, 3 dias ou 7 dias).

Returns

predictions: array_like

Array de tamanho 3, contendo as previsões de 1, 3 e 7 dias, respetivamente. Os valores vêm preenchidos a -1 sempre que não for pedido esse horizonte ou em caso de pedido inválido.

error_code : int

Inteiro indicador da validade dos inputs (ver tabela da função *validate_bp_prediction_data*).

Notes:

Dados introduzidos <u>inválidos</u> (verificado através da função *validate_bp_prediction_data*): *predictions* = [-1, -1, -1].

Dados introduzidos <u>válidos</u>: modelo de JNN previamente treinado efetua a previsão: independentemente do tamanho do *array* de entrada, *predictions* é um *array* de tamanho 3, sendo o primeiro valor a previsão correspondente ao primeiro dia após o input, o segundo valor a previsão correspondente ao terceiro dia após o input e o último valor a previsão correspondente ao sétimo dia após o input.

Examples:

```
>>> predictions, error_code = predict_bp([151,
98, '40'], None)
[-1, -1, -1], 202
>>> predictions, error_code = predict_bp([151,
98, 40], 3)
[0, 100.02143, 0], 0
```

validate_bp_prediction_data

validate_bp_prediction_data(patient_input, pred_horizon)

Avalia a validade dos inputs da função *predict_bp*, tanto o *array* de dados como a indicação do horizonte de previsão são avaliados. Valores fora da gama admissível são balizados e *missing values* são resolvidos, quando possível. Em casos de demasiados *missing values*, horizontes de predição inválidos, ou outros erros descritos na tabela, a função retorna o valor de ***error_code*** consoante o código de erro que discrimina a situação**.**

Parameters

patient_input: array_like

Array com dados de input. 1 valor de pressão arterial sistólica por dia. 3 a 10 valores, que devem ser inteiros ou *float.*

pred_horizon : int or None

O argumento do horizonte de previsão pode estar vazio (*None*) ou pode ser um número inteiro: 1, 3 ou 7 (correspondente a 1 dia, 3 dias ou 7 dias).

Returns

error_code : int

Inteiro indicador da validade dos inputs (ver **Tabela 3**).

patient_input: array_like

Array de entrada com as alterações efetuadas (caso aplicável).

Notes

O input é válido se cumprir todas as seguintes condições presentes na **Tabela 3**.

<table>
<tr><th>Parâmetro</th><th>error_code</th><th>Significado</th></tr>
<tr><td>-</td><td>0</td><td>Todos os dados introduzidos estão válidos.</td></tr>
<tr><td>pred_horizon</td><td>101</td><td>pred_horizon não corresponde a uma das opções válidas (None, 1, 3 ou 7).</td></tr>
<tr><td rowspan="3">patient_input</td><td>201</td><td>Os dados de entrada contêm NaN.</td></tr>
<tr><td>202</td><td>Tipos de valor inválidos nos dados de entrada (válidos: int e float).</td></tr>
<tr><td>203</td><td>Tamanho do input inferior a 3 ou superior a 10</td></tr>
</table>

Tabela 3 – Códigos de erro e respetivas descrições.

Examples

```
>>> error_code, patient_data =
validate_bp_prediction_data([151, 512, '40'],
None)
202, [151, 512, '40']
>>> error_code, patient_data =
validate_bp_prediction_data([151, 512, 40],
None)
0, [151, 200, 50]
```

B

Recommendation module API

DEFINIÇÃO DA INTERFACE DO MÓDULO DE RECOMENDAÇÃO DE PRESSÃO ARTERIAL

Listagem de métodos

- recommendations_bp
- validate_bp_recommendation_data

recommendations_bp

recommendations_bp (sex, age, BMI, waistCircumference, dateTime, SBP_DBP, heartRate, aerobicExerciseWeeklyDuration, aerobicExerciseWeeklyIntensity, resistanceExerciseWeekly, sodiumDaily, alcoholWeekly, comorbidities)

Retorna um *array* de tamanho 8. Quando os parâmetros são inválidos, o *array* atribui todos os valores -1. Quando os parâmetros se encontram válidos, o *array* é representado pelos valores 1, 2 e 3 (correspondentes a diminuir, manter e aumentar, respetivamente) em todas as posições, sendo as posições, por ordem, correspondentes às recomendações relativas a: índice de massa corporal (IMC), circunferência da cintura, duração semanal de exercício aeróbico, frequência semanal de exercício aeróbico, exercício de resistência semanal, ingestão diária de sódio, consumo semanal de álcool e consumo diário de álcool.

Parameters

sex: string
String referente ao sexo do utilizador. Deve ser inserido 'M' para masculino e 'F' para feminino.

age: int
Inteiro positivo maior que zero referente à idade do utilizador.

BMI: float
Float referente ao índice de massa corporal do utilizador, expresso em kg/m2.

waistCircumference: int
Inteiro referente à medição da circunferência de cintura do utilizador.

dateTime: string
String com informação acerca da data e hora da introdução das medições. Deve estar no formato *"dd-mm-YYYY HH:MM".*

SBP_DBP: string
String referente à medição da pressão arterial sistólica e diastólica do utilizador. Deve conter 2 inteiros, separados por '/', o primeiro referente à pressão sistólica e o segundo à pressão diastólica.

heartRate: int
Inteiro referente à medição da frequência cardíaca do paciente, em bpm (batimentos por minuto).

aerobicExerciseWeeklyDuration: array_like

Array composto por 7 inteiros maiores ou iguais que zero, em que cada posição corresponde à duração (min) de exercício físico aeróbico realizado em cada dia, para os últimos 7 dias.

aerobicExerciseWeeklyIntensity: array_like

Array composto por 7 inteiros, em que cada posição corresponde à intensidade do exercício físico aeróbico realizado em cada dia, para os últimos 7 dias. Os inteiros que constituem o *array* podem tomar três valores: 1 – moderada intensidade; 2 – alta intensidade; ou 0 – não foi efetuado exercício aeróbico.

resistanceExerciseWeekly: array_like

Array composto por 7 inteiros, em que cada posição corresponde à realização de exercício físico de resistência em cada dia, para os últimos 7 dias. Os inteiros que constituem o *array* podem tomar dois valores: 1 – realizou exercício físico de resistência; ou 0 – não foi efetuado exercício de resistência.

sodiumDaily: float

Float indicativo da quantidade de sódio ingerida no presente dia, em gramas.

alcoholWeekly: array_like

Array composto por 7 inteiros, em que cada posição corresponde ao consumo diário de álcool, para os últimos 7 dias. Cada elemento deve corresponder ao número de bebidas ingeridas (1 bebida = 125mL de vinho ou 250mL de cerveja).

comorbidities: string ou None

String referente às comorbidades do paciente. Caso não se aplique, pode estar vazio (*None*).

Returns

recommendations: array_like

Array composto por 8 inteiros, em que cada posição corresponde à recomendação resultante da avaliação dos parâmetros de input. As recomendações correspondem a, por ordem: IMC, circunferência da cintura, duração semanal de exercício aeróbico, frequência semanal de exercício aeróbico, exercício de resistência semanal, ingestão diária de sódio, consumo semanal de álcool e consumo diário de álcool.

error_code : int

Inteiro indicador da validade dos inputs (ver tabela da função *validate_bp_recommendation_data*).

Notes:

Condições dos parâmetros de input para cada regra (R01 – R12) são apresentados na **Tabela 5**. Alguns parâmetros de input não são apresentados nas tabelas porque não condicionam nenhuma regra. O parâmetro dailyAerobicDuration presente nas regras R05 a R06 resulta da multiplicação dos elementos em cada posição de aerobicExerciseWeeklyDuration com os respetivos elementos de aerobicExerciseWeeklyIntensity.

A **Tabela 6** apresenta os resultados das recomendações de cada parâmetro de saída conforme as condições dos parâmetros dos parâmetros de input para cada regra (R01 – R12), apresentados na tabela anterior.

Examples:

Dados introduzidos inválidos (verificado através da função *validate_bp_recommendation_data*): *recommendations* = [-1, -1, -1, -1, -1, -1, -1, -1].

Dados introduzidos válidos: independentemente dos dados de entrada, *recommendations* é um *array* de tamanho 8. Quando os parâmetros se encontram válidos, o *array* é representado pelos valores 1, 2 e 3 (correspondentes a diminuir, manter e aumentar determinada ação, respetivamente).

```
>>> recommendations, error_code =
recommendations_bp('F', 25, 18, 70, '16-08-
2022 17:42', '138/70', 60, [0, 0, 5, 15, 10,
25, 30], [0, 0, 2, 2, 2, 2, 1], [1, 1, 0, 0,
0, 0, 0], 1.5, [1, 2, 1, 1, 0, 0, 1], None)
[3, 2, 3, 2, 2, 2, 2, 1], 0
```

Rule ID	sex	BMI	waistCircumference	dailyAerobicDuration	resistanceExerciseWeekly	sodiumDaily	alcoholWeekly
R01		< 18.5					
R02		>= 25					
R03	M		>= 94				
R04	F		>= 80				
R05				sum(dailyAerobicDuration) < 150			
R06				dailyAerobicDuration .count(0) > 2			
R07					sum(resistanceExerciseWeekly) < 2		
R08						>= 2	
R09	M						sum(alcoholWeekly)[0] > 14
R10	F						sum(alcoholWeekly)[0] > 8
R11	M						any(x[0] > 2 for x in alcoholWeekly)
R12	F						any(x[0] > 1.5 for x in alcoholWeekly)

Tabela 5 - Condições dos parâmetros de input para cada regra (R01 – R12).

Rule ID	BMI	waistCircumference	aerobicExerciseWeeklyDuration	aerobicExerciseWeeklyFreq	resistanceExerciseWeekly	sodiumDaily	alcoholWeekly	alcoholDaily
R01	3							
R02	1							
R03		1						
R04		1						
R05			3					
R06				3				
R07					3			
R08						1		
R09							1	
R10							1	
R11								1
R12								1

Tabela 6 - Resultados das recomendações de cada parâmetro de saída conforme as condições dos parâmetros dos parâmetros de input para cada regra (R01 – R12).

validate_bp_recommendation_data

validate_bp_recommendation_data (sex, age, BMI, waistCircumference, dateTime, SBP_DBP, heartRate, aerobicExerciseWeeklyDuration, aerobicExerciseWeeklyIntensity, resistanceExerciseWeekly, sodiumDaily, alcoholWeekly, comorbidities)

Avalia a validade dos inputs da função *recommendations_bp*. Em casos de inputs inválidos, dependendo das condições de cada parâmetro, ou outros erros descritos na tabela, a função retorna o valor de ***error_code*** consoante o código de erro que discrimina a situação**.**

Parameters

sex: string
String referente ao sexo do utilizador. Deve ser inserido ‘M’ para masculino e ‘F’ para feminino.

age: int
Inteiro positivo maior que zero referente à idade do utilizador.

BMI: float
Float referente ao índice de massa corporal do utilizador, expresso em kg/m2.

waistCircumference: int
Inteiro referente à medição da circunferência de cintura do utilizador.

dateTime: string
String com informação acerca da data e hora da introdução das medições. Deve estar no formato *"dd-mm-YYYY HH:MM".*

SBP_DBP: string
String referente à medição da pressão arterial sistólica e diastólica do utilizador. Deve conter 2 inteiros, separados por ‘/’, o primeiro referente à pressão sistólica e o segundo à pressão diastólica.

heartRate: int
Inteiro referente à medição da frequência cardíaca do paciente, em bpm (batimentos por minuto).

aerobicExerciseWeeklyDuration: array_like
Array composto por 7 inteiros maiores ou iguais que zero, em que cada posição corresponde à duração (min) de exercício físico aeróbico realizado em cada dia, para os últimos 7 dias.

aerobicExerciseWeeklyIntensity: array_like

Array composto por 7 inteiros, em que cada posição corresponde à intensidade do exercício físico aeróbico realizado em cada dia, para os últimos 7 dias. Os inteiros que constituem o *array* podem tomar três valores: 1 – moderada intensidade; 2 – alta intensidade; ou 0 – não foi efetuado exercício aeróbico.

resistanceExerciseWeekly: array_like

Array composto por 7 inteiros, em que cada posição corresponde à realização de exercício físico de resistência em cada dia, para os últimos 7 dias. Os inteiros que constituem o *array* podem tomar dois valores: 1 – realizou exercício físico de resistência; ou 0 – não foi efetuado exercício de resistência.

sodiumDaily: float

Float indicativo da quantidade de sódio ingerida no presente dia, em gramas.

alcoholWeekly: array_like

Array composto por 7 inteiros, em que cada posição corresponde ao consumo diário de álcool, para os últimos 7 dias. Cada elemento deve corresponder ao número de bebidas ingeridas (1 bebida = 125mL de vinho ou 250mL de cerveja).

comorbidities: string ou None

String referente às comorbidades do paciente. Caso não se aplique, pode estar vazio (*None*).

Returns

error_code : int

Inteiro indicador da validade dos inputs (ver **Tabela 9**).

bp_category: string

String referente à categoria de pressão arterial atribuída consoante os níveis de pressão arterial do utilizador, segundo a classificação presente em [1]. Pode ser: 'OPTIMAL', 'NORMAL', 'HIGH NORMAL', 'GRADE 1 HYPERTENSION', 'GRADE 2 HYPERTENSION', 'GRADE 3 HYPERTENSION'. Caso exista algum erro nos inputs, devolve -1.

sbp: int

Inteiro referente à pressão arterial sistólica do utilizador, aferido a partir do parâmetro de input *sbp_dbp*. Caso exista algum erro no input, devolve -1.

dbp: int

Inteiro referente à pressão arterial diastólica do utilizador, aferido a partir do parâmetro de input *sbp_dbp*.

dateTime: timestamp

Timestamp referente à data e hora da introdução das informações do utilizador, aferido a partir da string de input *dateTime*.

Notes

O input é válido se cumprir todas as seguintes condições presentes na **tabela 9**.

Parâmetro	*error_code*	Significado
-	0	Todos os dados introduzidos estão válidos.
sex	10	sex deve ser uma string: ‘M’ – Masculino ou ‘F’ - Feminino
age	20	age deve ser um inteiro maior que 0
BMI	31	BMI deve ser um float maior que 12.0
waistCircumference	32	waistCircumference deve ser um float
dateTime	40	dateTime deve ser uma string que indica a data e hora no formato 'dd/mm/yyyy HH:MM'
SBP_DBP	50	SBP_DBP deve ser uma string composta por 2 inteiros separados por ‘/’, que indicam o valor da medição de pressão sistólica e diastólica, respetivamente ('SBP/DBP' em mmHg)
heartRate	60	heartRate deve ser um inteiro positive, relative à medição da frequência cardíaca (bpm)
aerobicExerciseWeeklyDuration	70	aerobicExerciseWeeklyDuration deve ser list ou array composta por 7 inteiros maior ou iguais a 0, com a duração diária (min) de exercício aeróbico ao longo da última semana
aerobicExerciseWeeklyIntensity	71	aerobicExerciseWeeklyIntensity deve ser list ou array composta por 7 inteiros, com a intensidade diária (1 - moderada, 2 - alta ou 0 – não realizado) de exercício aeróbico ao longo da última semana
resistanceExerciseWeekly	72	resistanceExerciseWeekly deve ser list ou array composta por 7 inteiros, com a realização diária (1 - realizado, 0 - não realizado) de exercício de resistência ao longo da última semana
sodiumDaily	80	sodiumDaily deve ser um float maior ou igual que 0, indicative do consumo de sódio nesse dia (g)
alcoholWeekly	90	alcoholWeekly deve ser list ou array composta por 7 inteiros, com o consumo diário de álcool ao longo da última semana, em unidades de bebidas (1 unidade = 125 mL de cerveja ou 250 mL de vinho, que corresponde a cerca de 8g de álcool)

comorbidities	100	comorbidities string com as doenças existentes do paciente, separadas por virgula, ou None, se não se aplicar (‘hipertensão, diabetes, obesidade, dislipidemia’)

Tabela 9 – Códigos de erro e respetivas descrições.

Examples

```
>>> validate_bp_recommendation_data ('F', 25, 18, 70, '16-08-2022 17:42', '138/70', 60, [0, 0, 5, 15, 10, 25, 30], [0, 0, 2, 2, 2, 2, 1], [1, 1, 0, 0, 0, 0, 0], 1.5, [1, 2, 1, 1, 0, 0, 1], None)
0, 'HIGH NORMAL', 138, 70, datetime.datetime(2022, 8, 16, 17, 42)
```

[1] Williams, B., Mancia, G., Spiering, W., Agabiti Rosei, E., Azizi, M., Burnier, M., … Dominiczak, A. (2018). 2018 ESC/ESH Guidelines for the management of arterial hypertension. European Heart Journal. doi:10.1093/eurheartj/ehy33